THE

LONDON, EDINBURGH, AND DUBLIN

PHILOSOPHICAL MAGAZINE

AND

JOURNAL OF SCIENCE.

[SEVENTH SERIES.]

JULY 1933.

I. *The Ellipsoidal Viscosity Distribution.* *By* W. S. KIMBALL, *Ph.D., Michigan State College, East Lansing, Michigan* *.

INTRODUCTION.

DISTRIBUTION functions for viscosity and heat conduction have been obtained by Chapman † and Enskog ‡ using successive approximations in solving the transport equation and Boltzmann's continuity equation. The present treatment employs the geometrical weight § method that deals with strains and stresses in velocity and ordinary space, and accordingly every conclusion is based on Newtonian mechanics and elasticity theory, and probability theory is just incidental. This involves the taking over into velocity space of those parts of elasticity theory that relate to homogeneous strain, simple shear, and the strain ellipsoid||.

The accuracy is here of the same order as that of the gas law and Maxwell's velocity distribution. Refinements such as the Pauli exclusion principle, Van der Waals's equation, the new statistics, and the nature of

* Communicated by the Author.

† S. Chapman, Phil. Trans. Roy. Soc. A, p. 216 (1916).

‡ Enskog, Inaugural Dissertation, Upsala (1917).

§ W. S. Kimball, Jr. Phys. Chem. xxxv. p. 611 (1931).

|| A. G. Webster, 'Dynamics,' chap. ix.

collisions are not considered, and large scale elastic stresses replace any particular law of force between molecules.

On the other hand, new contributions are involved in that, first, the present distribution provides an exact solution of Boltzmann's equation. Secondly, it purports to show that no single distribution can adequately portray viscosity in gas unless it includes a parameter that varies with position along the molecular free paths. This is because the distribution of velocities immediately before an equivalent impact appears to be fundamentally different from the distribution immediately after impact (being its mechanical image), as well as from the series of distributions between these extremes. In this connexion the free path concept is introduced, but *mean* free paths do not enter the distribution formula to impair the rigour of the theory.

§ 1. *The Viscosity Shear Law : the mean impressed Shear per Impact is constant throughout Velocity Space.*

We consider an xyz coordinate system in gas having a mass flow or drift velocity u_0 in the x direction whose magnitude varies with the height z only, its gradient being a negative constant. Thus if the origin moves along with the gas, $u_0 = -Bz$, indicating a negative drift velocity above the xy plane and positive below it, $\frac{du_0}{dz} = -B$ is the drift of the gas in the x direction, per unit time per unit vertical distance along z, *i. e.*, the time rate of shear in the negative x direction. Multiply by $t_1 = \frac{\lambda}{\bar{c}} = \frac{1}{n_1}$, the average time interval between impacts or reciprocal of the number of impacts per second, λ being the mean free path, and $\bar{c}$ the mean velocity :

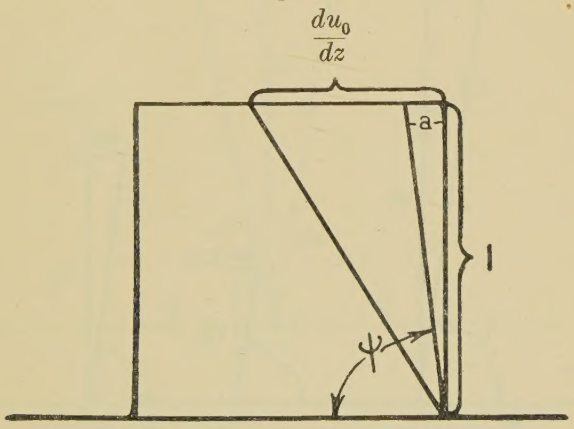
$$\left.\begin{aligned} a = -t_1 \frac{du_0}{dz} = -\frac{\lambda}{\bar{c}} \frac{du_0}{dz} = \cot \psi, \\ \text{or} \qquad -\frac{du_0}{dz} = \frac{\bar{c}}{\lambda} a = n_1 a. \end{aligned}\right\} \quad . \quad . \quad (1)$$

Fig. 1 shows how the shear of the gas per unit time $\frac{du_0}{dz}$ is n_1 times the separate amounts of shear for the times $t_1 = \frac{\lambda}{\bar{c}}$. But each molecule has in the mean

a single impact during this time, and the amount of shear involved must be the same for each molecule, since (1) refers to the homogeneous shearing of the gas that includes them all. Thus, when averaged in this way, the mean shear per impact of all molecules fast and slow must be alike at a given place, for otherwise they would be separated into fast and slow groups along the line of gas flow. And this does not occur during the time t_1 or any other time interval, for then the existing homogeneity of horizontal layers of gas would not obtain. Hence the viscosity shear law as above stated.

This conclusion may be checked by consideration of the impressed force that causes the rate of shear in the gas.

Fig. 1.



The time rate of shear and the mean impact shear.

Let F′ be its amount per unit mass at (u, v, w) in velocity space. Then, by Newton's second law,

$$\left.\begin{aligned} \mathrm{F}' = \frac{du_0}{dt} &= u\frac{\partial u_0}{\partial x} + v\frac{\partial u_0}{\partial y} + w\frac{\partial u_0}{\partial z} = w\frac{du_0}{dz} \\ &= -aw\frac{c}{\lambda}, \end{aligned}\right\} \quad . \quad . \quad (2)$$

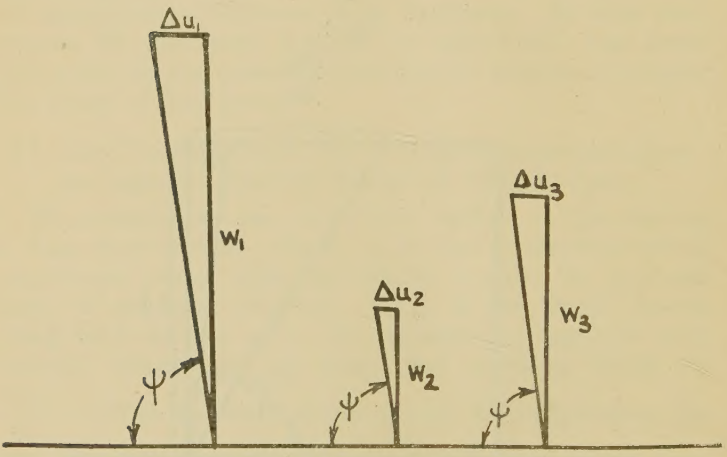
or

$$\left.\begin{aligned} m\mathrm{F}' = mw\frac{du_0}{dz} &= -maw\frac{c}{\lambda} = m\Delta u\frac{c}{\lambda}, \\ &\text{per molecule of mass } m. \end{aligned}\right\}$$

Eqs. (2) are a matter of definition and Newtonian Mechanics, and it is noteworthy that F′ which, by definition, acts only in the x direction is proportional to w. In the right members we have (using (1)) introduced $\bar{c}/\lambda$, the average number of impacts per second, and we note that the other factors, being the changes of momentum per impact, are required by (2) to be proportional to w. And this leads again to the viscosity shear law above stated :

$$\Delta u = -aw\,; \quad -\frac{\Delta u}{w} = a = \cot\psi, \quad . \quad . \quad . \quad . \quad (3)$$

Fig. 2.



The equal impressed shears for various velocities.

whereby the mean amount of shear per impact is the same for all speeds, as shown in fig. 2.

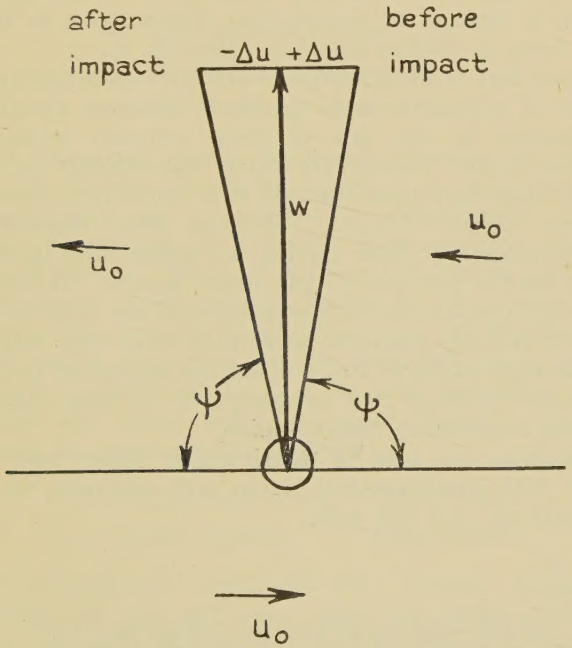
By elementary theory $1/\lambda$ is the number of impacts in unit distance travelled. Then if a is the constant mean shear per impact, a/λ is the amount of shear per unit distance travelled for all speeds and directions. If all molecules had the same speed c then $c\ a/\lambda$ would be the rate of shear (1) of them all. Since each one has $\bar{c}$ for its average speed, the latter appears in (1) instead of c.

Another check upon (1) is given at the end of this paper, where eq. (90) is derived, giving the relation between $\frac{du}{dz}_0$ and the amount of shear a.

§ 2. *Velocity Distributions before and after Impact.*

In case of the familiar Maxwellian distribution of spherical symmetry, no attempt is usually made to distinguish between molecular behaviour before, after, or between impacts. And the obvious reason for this is that, generally speaking, molecular behaviour is exactly the same before, after, and between impacts, in respect

Fig. 3.



The mean shear before and after impact.

to probability of velocity and momentum magnitude, direction, etc.

For the viscosity distribution under discussion, however, it is clear that velocities before and after impact will not be the same in respect to probability of magnitude and direction. Thus we have seen, as shown by fig. 2, that a negative increment $\Delta u = -aw$ is *imparted* upon impact to an average molecule moving vertically with velocity w. By Newton's third law, however, it must have an equal and opposite positive increment $\Delta u = aw$

before impact if the assumed steady state is to be maintained. Now this positive increment aw is easily seen to result from the kinematics of the situation due ultimately to the impressed force. For a molecule coming from below where the gas moves forward will have, in all probability, a positive x increment of velocity and *vice-versa* for molecules coming from above. This positive increment has to be wiped out upon impact and reversed in direction with equal magnitude according to fig. 3 and the viscosity shear law. This process involves a change of momentum which goes to make up the force of viscous reaction. (See fig. 8.)

As used here the term impacts always means equivalent impacts of negligible time duration, without specifying their nature or the law of force between molecules. Velocities before impact are those *immediately* before—so close that zero-time interval and zero-space interval intervene before impact. Velocities after impact are those *immediately* after having an *entire* free path for space interval previous to the next impact. It is these *latter* velocities whose average *determines* the *drift* velocity and drift velocity gradient, since they have time and free path through which to act, and are not expunged immediately upon coming into existence, as it were, like the velocities *immediately before* impact.

To express the idea of distributions before and after impact with mathematical rigour and clearness, we may anticipate eq. (81) and write

$$dn = \frac{n}{\alpha^3 \pi^{3/2}} \exp\left[-\left(\frac{u^2+v^2+w^2+a^2w^2(2x-1)^2-2auw(2x-1)}{\alpha^2}\right)\right]_{du\,dv\,dw\,dx} \quad . \quad . \quad . \quad (81\,a)$$

for the number (out of the n per c.c.) of molecules within the element $du\,dv\,dw$ of velocity space, and which also have traversed a fraction between x and $x+dx$ of their several total free paths. Thus x is the dimensionless fraction of the free path traversed since last previous impact and ranges from 0 to 1. When x equals 1 or 0, (81 a) gives the distribution (21) before and after impact respectively, and when x equals 1/2, it gives the Maxwell distribution of spherical symmetry which we shall

see obtains for molecules at the mid-points of their free paths.

§ 3. *The Strain Ellipsoids in Velocity Space.*

The viscosity shear law (3) makes it possible to employ in velocity space part of the theory of elasticity. Thus *homogeneous strain* is represented by linear relations between the relative displacements and position coordinates of points in velocity space :

$$\left.\begin{aligned} \Delta u &= a_1 u + a_2 v + a_3 w, \\ \Delta v &= b_1 u + b_2 v + b_3 w, \\ \Delta w &= c_1 u + c_2 v + c_3 w. \end{aligned}\right\} \quad . \quad . \quad . \quad . \quad . \quad (4)$$

The diagonal terms represent *stretches* or *squeezes* and the others represent *shears*. A *pure strain* is characterized by a symmetric determinant in (4). When (4) reduces to a single diagonal term we have a *simple strain*. When it reduces to a single skew term we have a *simple shear*. Evidently this latter term describes the case in hand, since eqs. (3), together with figs. 2 and 3 show that all the constants of (4) are zero except a_3, which equals $+a$ or $-a$ according as we consider the shear before or after impact. Thus (4) becomes the viscosity law of shear :

$$\begin{aligned} \Delta u = aw \quad &\text{or} \quad \Delta u = -aw. \quad . \quad . \quad (5) \\ \text{before impact} \quad &\text{after impact,} \end{aligned}$$

which completely accounts for viscous phenomena according to (2) and (3). Eq. (5) forms the basis of two complete geometrical structures in velocity space describing the situations *before* and *after* impact. The constant $a = \cot \psi$ involved in (5) thus appears in a familiar role. It is "commonly called the *amount* * of shear." According to elasticity theory, points originally on a sphere will lie after a displacement on a strain ellipsoid at positions determined by the transformation eqs. (4) for homogeneous strain. We use a polar coordinate system in velocity space as well as cartesian coordinates taking the y axis as the polar

* A. G. Webster, 'Dynamics,' chap. ix. p. 441.

axis, and the origin moving along with and hence at rest in the gas.

$$\left.\begin{aligned} u &= c \sin\theta \cos\phi, \\ v &= c \cos\theta, \\ w &= c \sin\theta \sin\phi, \end{aligned}\right\} \quad \ldots\ldots \quad (6)$$

and using the primed variables to locate points on a sphere, we have the transformation :

$$\left.\begin{aligned} u' = c' \sin\theta' \cos\phi' &= c \sin\theta \cos\phi \\ &\quad \pm ac \sin\theta \sin\phi = u \pm aw, \\ v' = c' \cos\theta' &= c \cos\theta = v, \\ w' = c' \sin\theta' \sin\phi' &= c \sin\theta \sin\phi = w. \end{aligned}\right\} \quad \ldots \quad (7)$$

Hence the y and z components of velocity are seen to be unchanged, since (4) and (5) provide for no changes in v and w and, according to the present theory, the force and strain equations (2) and (3) account completely for the facts of viscosity in elementary kinetic theory. The minus and plus signs in the first of eq. (7) refer to the situation before and after impacts respectively, showing u greater than u' by $\Delta u = aw$ or *vice versa* according to (5). The y and z components in (7) refer to components either before or after impact. We square and add eqs. (7) to get the strain quadric :

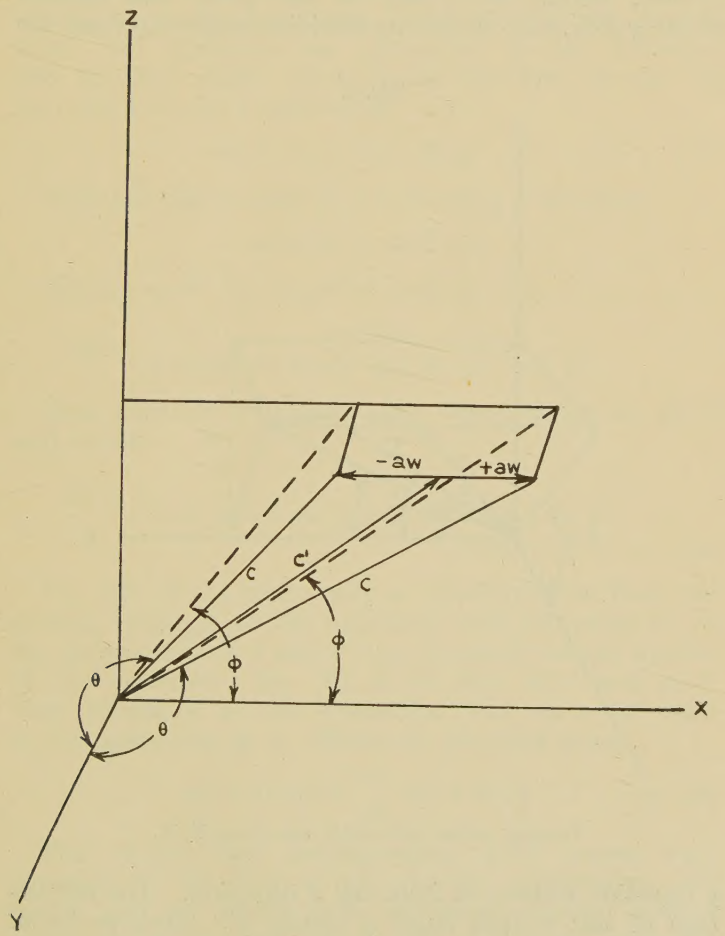
$$\left.\begin{aligned} c'^2 &= c^2(1 + a^2 \sin^2\theta \sin^2\phi \pm 2a \sin^2\theta \sin\phi \cos\phi) \\ &= c^2 + a^2w^2 \pm 2\, caw \cos \mathrm{Q}, \\ \text{or} \qquad c'^2 &= u^2 + v^2 + w^2 + a^2w^2 \pm 2auw. \end{aligned}\right\} \quad . \quad (8)$$

Here the right member of the polar equation shows the relation between c, c', and aw as the law of cosines, and fig. 4 shows that $c \cos\phi = u = c \sin\theta \cos\phi$. The two sets of coordinates (c, θ, ϕ) shown in fig. 4 represent velocities *before* and *after* impact, the latter showing the shift $-aw$ to the left in the direction of drift velocity that exists *above* the xy plane.

The strain quadrics given by (8) are evidently strain ellipsoids in velocity space. Each point of their surfaces represents the direction and magnitude of velocities c existing under viscosity conditions *before* and *after* impact, determined by the transformations (7), which give the elastic displacements which distort the spherical

symmetry of distribution according to the law of shear (5). There is, of course, an infinity of such strain ellipsoids (8) corresponding under (7) to the infinity of spheres

Fig. 4.



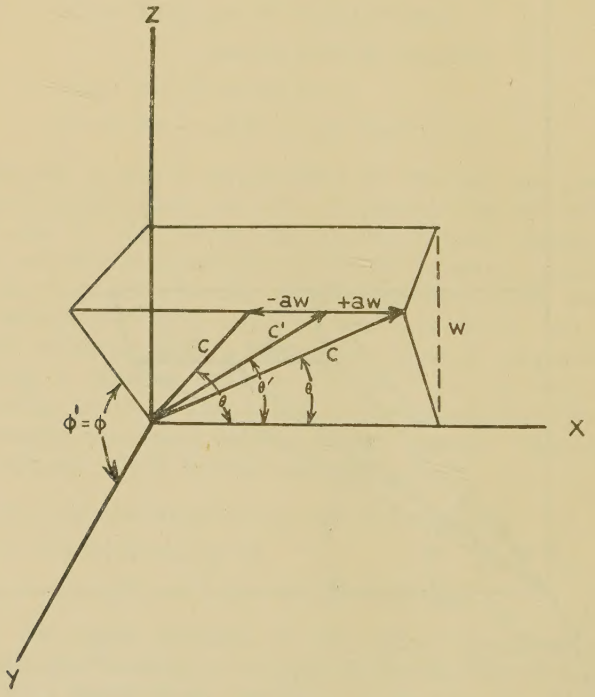
Velocity strains with polar axis along Y, Y.

having velocity magnitudes given by the parameter c'. We show below that these spheres, which are subjected to no shear of the type (5), belong to a spherical distribution of the Maxwell type for the same temperature.

§ 4. *The Solid Angles that include Equal Numbers of Points on the Strain Ellipsoids vary inversely as the Cube of the Velocity Magnitude.*

Now refer our ellipsoid to a different polar coordinate system, taking the x axis as the polar axis instead of the y axis, and measuring meridian angles up from the

Fig. 5.



Velocity strains with polar axis along X, X.

y direction instead of from the x direction. The advantage of this is that angle ϕ equals ϕ'. Then we have, instead of (7),

$$\left.\begin{aligned} u' &= c' \cos\theta' = c\cos\theta \pm ac \sin\theta \sin\phi = u \pm aw, \\ v' &= c' \sin\theta' \cos\phi' = c \sin\theta \cos\phi = v, \\ w' &= c' \sin\theta' \sin\phi' = c \sin\theta \sin\phi = w, \\ & c' \sin\theta' = c \sin\theta. \end{aligned}\right\} \quad . \quad . \quad (9)$$

Square and add as before, and obtain

$$c'^2 = c^2(1+a^2 \sin^2\theta \sin^2\phi \pm 2a \sin\theta \cos\theta \sin\phi) \quad . \quad . \quad (10)$$

This is exactly the same ellipsoid as (8) expressed in different coordinates. Comparison of the last two of eqs. (9) shows that angles ϕ and ϕ' are equal, so that these two equations are equivalent to the indicated one on their right. Now divide the first of eqs. (9) by $c' \sin\theta' = c \sin\theta$ and obtain

$$\cot\theta' = \cot\theta \pm a \sin\phi \quad . \quad . \quad . \quad . \quad . \quad (11)$$

Take the differentials of (11), keeping ϕ constant :

$$csc^2\theta' d\theta' = csc^2\theta \, d\theta,$$

Replace $csc^2\theta'$ by $1+\cot^2\theta'$ and use (11) :

$$d\theta' = \frac{csc^2\theta \, d\theta}{(1+\cot^2\theta + a^2 \sin^2\phi \pm 2a \cot\theta \sin\phi)}.$$

Now multiply numerator and denominator by $\sin^2\theta$ and obtain

$$d\theta' = \frac{d\theta}{(1+a^2 \sin^2\theta \sin^2\phi \pm 2a \sin\theta \cos\theta \sin\phi)}$$

$$= d\theta(c^2/c'^2) \quad . \quad . \quad (12)$$

Eq. (12) shows that under a transformation from the primed variables to the unprimed ones according to (9) the differentials of θ and θ' vary inversely as the squares of the velocities. Eq. (9) shows that $\sin\theta$ and $\sin\theta'$ vary inversely as the velocities, and $d\phi = d\phi'$ since no changes occur in ϕ. Hence for the solid angles,

$$\sin\theta' d\theta' d\phi' = \frac{c^3}{c'^3} \sin\theta \, d\theta \, d\phi \quad . \quad . \quad . \quad (13)$$

This shows that corresponding solid angles vary inversely as the cube of the velocity magnitudes. All points and no others that terminate velocity vectors within the solid angle on the left of (13) will, after the transformation (9) terminate within the solid angle on the right of (13), there being under (9) a one to one correspondence between points on the sphere and on the ellipsoid. The density of points on the sphere is uniform, since it refers to a Maxwell distribution with no distorting shear and corresponding to a constant value

of c', and hence equal solid angles for the sphere (left number of (13)) include equal numbers of points on the sphere representing molecular velocities, and these same equal numbers of corresponding points on the ellipsoid are included according to (13) by unequal solid angles, varying inversely as c^3.

§ 5. *The Simple Shear depending on the Single Constant "a" involves Distortion but no Expansion, and hence the Volume Density of Points in Velocity Space between successive Ellipsoids is uniform and equal to that between the corresponding Spheres of the Maxwell Distribution.*

From (8) or (10) note that, for changes in c' and c subject to constant θ and ϕ, we have

$$dc' = \frac{c'dc}{c};$$

and hence

$$c'^2\,dc' = \left(\frac{c'}{c}\right)^3 c^2\,dc \quad . \quad . \quad . \quad . \quad . \quad . \quad (14)$$

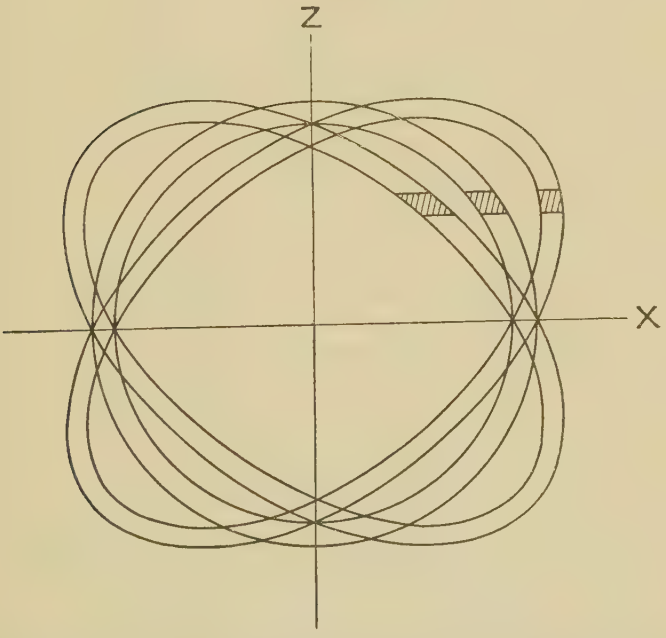
The product of (13) by (14) gives for the volume elements

$$c'^2 \sin\theta' dc' d\theta' d\phi' = c^2 \sin\theta\, dc\, d\theta\, d\phi \quad . \quad . \quad . \quad . \quad (15)$$

Eq. (15) shows that any volume element, bounded by two successive nearby spheres of the spherical distribution, and an including solid angle, is equal to the volume element that corresponds to it under transformation (9) between successive ellipsoids and the corresponding solid angles. The spherical distribution, however, has uniform density between successive spheres for points in velocity space representing velocity direction and magnitude. All of these points included in such a volume element will under (9) be carried over to corresponding points included in the corresponding volume element included between successive ellipsoids, the latter volume element being by (15) the same size as the former. Hence the density or concentration of points in velocity space between the ellipsoids is the same as between the spheres of the Maxwell distribution for each pair of corresponding volume elements, and,

being uniform all around the spherical distribution, the density will also be uniform throughout the corresponding space between the ellipsoids. Fig. 6 shows the xz section of the space included between successive strain ellipsoids before and after impact that correspond under (9). Each includes the same number of points and uniformly distributed and hence with equal density

Fig. 6.



Sections showing distorted but unexpanded velocity space of uniform molecular density between successive ellipsoids and spheres.

for the two cases, thus showing distortion in velocity space but no expansion.

§6. *Criticism of the Viscosity Shear Law.*

It is worth noting that the shear law (5) gives a very simple, perhaps the simplest possible, departure from spherical symmetry in distribution that is consistent with obvious facts of viscosity. The y and z velocity

components are unaffected, since temperature is assumed uniform, and we consider no viscous force except in the x direction within the gas. Furthermore, the correction for the x component, Δu, will obviously depend primarily on the z component, w, which is in the direction of the drift velocity gradient; for, if the molecule comes from below the xy plane, it comes from gas with a forward drift velocity with respect to the origin of coordinates, and will in all probability have a positive increment added to its x component by the time impact takes place, and *vice versa* immediately after impact. In general $\Delta u = f(u, v, w)$, and if this be expanded in powers of u, v, and w by Maclaurin's series, the first term of the expansion will be $\Delta u = f'(0)w = \pm aw$. (The constant term and powers of u and v only vanish since $\Delta u = 0$ when $w = 0$.) Thus the viscosity shear law above represents a first approximation which, nevertheless, is very accurate, as we shall see below, and hence might be legitimately used herein even as a hypothesis. We show by (96) below that "a" is usually smaller than 10^{-6}, and higher power terms in the expansion of Δu are probably entirely negligible.

§7. *Standard Form of the Ellipsoid in Cartesian Coordinates.*

To find the direction of maximum c, note that the ellipsoids (8) are symmetric with respect to the plane $v = 0$ since v appears only raised to the second power. Therefore the maximum of c and the major axis of shear will lie in this plane, and we take accordingly $\theta = \pi/2$ and, keeping c' constant, set the derivative of c^2 with respect to ϕ equal to zero. Thus the derivative of $c^2 = c'^2/(1 + a^2 \sin^2 \phi \pm 2a \sin \phi \cos \phi)$ vanishes when $\pm 2 \cot 2\phi = a$, and letting the constant b represent this critical value of ϕ, we have

$$\pm a = 2 \cot 2b = \cot b - \tan b \quad . \quad . \quad . \quad (16)$$

for the relation between amount of shear a and the angle b which the major axis of shear makes with the x axis or direction of viscous flow. Now take the original ellipsoid (8) and rotate the coordinate system about the polar axis (the y axis) through the angle b, so that the new primed coordinate axes will be axes of symmetry

of the ellipsoid ; use the cartesian expression (8) for the ellipsoid, and the transformation equations

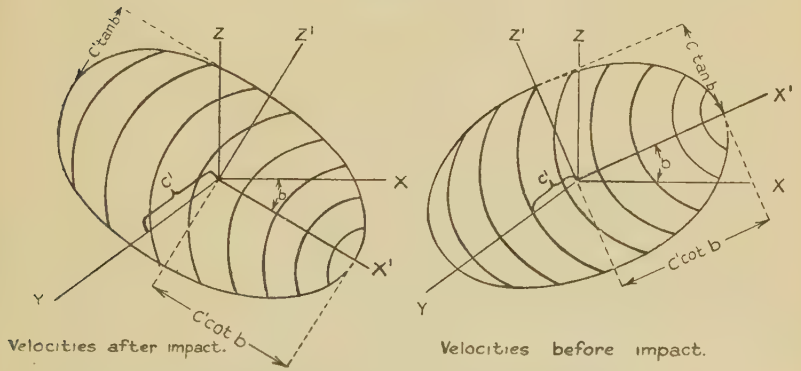
$$\left.\begin{aligned} u &= x' \cos b \pm z' \sin b, \\ v &= y', \\ w &= \pm x' \sin b + z' \cos b, \end{aligned}\right\} \quad . \quad . \quad . \quad (17)$$

and then, after a bit of complicated but elementary trigonometry and taking account of (16), we obtain

$$\frac{x'^2}{(c' \cot b)^2} + \frac{y'^2}{c'^2} + \frac{z'^2}{(c' \tan b)^2} = 1 \quad . \quad . \quad . \quad (18)$$

for both the ellipsoids (8) referred to their own axes of symmetry.

Fig. 7.



The strain ellipsoids before and after impact.

We note from (16) that the range of b is from $\pm\pi/4$, where $a=0$ and the viscous flow is zero, to $b=0$, where $a=\infty$ and the viscous flow is infinitely fast. Actually the departure on the part of b from $\pm\pi/4$ is very small, since the u components of velocity are only slightly larger than the w components. (Viscous flux velocity is small compared with agitation velocity magnitudes and "a" turns out to be less than 10^{-6}, and $\pi/4-b$ less than 1 per cent. of a second of arc, and the eccentricities of (18) are less than 10^{-3}.)

We may see clearly why the major axes of the ellipsoids lie nearest the x axis so that b is between $\pm\pi/4$

and zero. It is because the x components of velocity are all increased over their value in a spherically symmetric distribution, by a plus and minus amount before and after impact respectively, whereas none of the z components are affected, and hence the axis of symmetry of the ellipsoid will be turned from the mid-point between the x and z axes and rotated slightly towards the x axis, but never actually reaching it until the z components of velocity are negligible compared with the x components.

§ 8. *The Standard Form of the Ellipsoid in Polar Coordinates.*

The most convenient form for the ellipsoids is when they are referred to the coordinate system (6), using, however, b as the viscosity parameter instead of a. This form is readily obtained from (18) under the transformation

$$\left.\begin{aligned} x' &= c \sin\theta \cos(\phi \pm b), \\ y' &= c \cos\phi, \\ z' &= c \sin\theta \sin(\phi \pm b). \end{aligned}\right\} \quad . \quad . \quad . \quad . \quad (19)$$

It is

$$\left.\begin{aligned} c^2 &= \frac{c'^2 \tan^2 b}{(1-(1-\tan^2 b)\cos^2\theta - (1-\tan^4 b)\sin^2\theta\cos^2(\phi \pm b)} \\ &= \frac{c'^2 \tan^2 b}{(1-p\cos^2\theta - q\sin^2\theta\cos^2(\phi \pm b)}, \\ \text{where} \quad & p = 1-\tan^2 b = e^2_1, \; q = 1-\tan^4 b = e_2{}^2. \end{aligned}\right\} . \quad (20)$$

Eq. (20) gives not only the same identical ellipsoid as (8), but it is expressed in terms of the same identical coordinates as (8). This is proved by substituting (19) in (17) to get (6). The difference in form results from using the parameter b instead of a. Hence (8) and (20) can be transformed into each other with the help solely of (16) and some elementary though intricate trigonometry.

It is noteworthy that p and q are each always positive, since angle b is less in magnitude than $\pi/4$. Furthermore, $\sqrt{p}$ is the eccentricity of the elliptical trace of our ellipsoid in the yz' plane, whose equation is obtained by setting $\phi = \pi/2 \pm b$. This reduced equation is the standard form in polar coordinates for an ellipse with semi-minor axis equal to $c' \tan b$ and with origin at the centre of ellipse. Likewise, the trace of our ellipsoid in the plane xz

(same as $x'z'$ plane) is obtained by setting $\theta=\pi/2$, and is seen to be also the standard form in polar coordinates of an ellipse with centre at the origin and semi-minor axis equal to $c'\tan b$, but this time the eccentricity is $\sqrt{q}$. Our ellipsoids are, of course, not surfaces of revolution, but each is a reflexion or mechanical image of the other in the xy or yz plane. For small (actual) values of a, $e_2=\sqrt{2}e_1$, as appears from (16) and (20). These eccentricities are usually less than 10^{-3} (see section 22 below).

§ 9. *The Strain Ellipsoids are Equal Probability Loci in Velocity Space.*

We now proceed to set up our distribution function and evaluate its constants and show that it meets all the requirements of the case in hand. We have

$$dn=nf\,d\omega=n\mathrm{A}\exp\left[\frac{-c'^2}{a^2}\right]d\omega=n\mathrm{A}\exp\left[\frac{-(c^2(1-p\cos^2\theta-q\sin^2\theta\cos^2(\phi\pm b))}{a^2\tan^2 b}\right]c^2\sin\theta\,d\theta\,d\phi\,dc. \quad . \quad (21)$$

for the number of molecules in the element $d\omega=c^2\sin\theta\,d\theta\,d\phi\,dc$ of velocity space before or after impact according as we use minus or plus b in (21). Here the right member is evidently obtained by replacing c' by its value according to (20), and the constants A and a are to be determined below.

Divide by n :

$$\frac{dn}{n}=\mathrm{A}\exp\left[\frac{-c^2(1-p\cos^2\theta-q\sin^2\theta\cos^2(\phi\pm b))}{a^2\tan^2 b}\right]_{d\omega} \quad (22)$$

This equation shows that the exponential term on the right is constant as c varies according to (20), keeping c' constant, *i. e.*, the left member, or probability, is constant over the strain ellipsoid for a given volume element $d\omega$. This proves that (8) and (20) are equal probability ellipsoids, and incidentally that (5), which connects them with the equal probability spheres of the Maxwell distribution, can be viewed as a *probability* law or hypothesis. Furthermore, the constancy of the exponential term as c varies over the ellipsoid with c'

constant shows by (21) that under this condition dn is proportional to the volume element $d\omega$, so that the distribution (21) shows distortion under shearing stress but no expansion, *i. e.*, the density of points representing molecules in velocity space between successive ellipsoids dc apart is uniform in agreement with § 5. Note also that the spheres referred to by the primed variables and related by the law of shear (5) to the shear ellipsoids according to (7) and (9) now appear in (21), with c'^2 as spheres of a Maxwell or spherical distribution of the c''s.

§10. *The Ellipsoidal Viscosity Distribution Functions satisfy the Equilibrium Conditions for Stress and Strain acting in Velocity Space, that determine Maximum Entropy or Strain in each Direction.*

By reliance on the geometrical expression for weight,

$$W=n^n(r_1r_2r_3\ldots.r_n)=\frac{I}{f(u_1)f(u_2)\ldots.f(u_n)}, \quad . \quad (23)$$

it has been shown * that the equilibrium equations which, by the Lagrange method, determine the state of maximum entropy, strain, and weight, are equilibrium equations between stress and strain acting in velocity space; thus the equations for the x component of velocity in the spherically symmetric case

$$-W\frac{d}{du_i}(\log f(u_i))+\lambda mu_i=W\frac{d}{du_i}(\log r_i)+\lambda mu_i=0 \quad . \quad (24)$$

are the same as (replace the undetermined constant λ by its value $-W/kT$, found in the usual way)

$$kT=\frac{mu_1du_1}{dr_1/r_1}=\frac{mu_2du_2}{dr_2/r_2}\ldots\ldots=\frac{mu_ndu_n}{dr_n/r_n} \quad . \quad (25)$$

showing that kT is the modulus of elasticity per molecule for this case.

We may apply the same procedure to (21), considering separately *each direction* at a time. As the n molecules per c.c. make impacts, the number of these with direction

* Kimball, *loc. cit.*

determined by θ and ϕ in unit volume element of velocity space is

$$\frac{dn}{d\omega} = nf(c) \text{ ; where } f(c) = \mathrm{A} \exp \left[\frac{-c^2(1-p\cos^2\theta - q\sin^2\theta\cos^2(\phi \pm b))}{a^2\tan^2 b} \right]. \quad . \quad (26)$$

Hence the range, or volume in velocity space occupied per molecule, is

$$r = \frac{1}{nf(c)}, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (27)$$

and the corresponding weight per molecule is

$$w = \frac{1}{f(c)} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (28)$$

and the weight for all of the comparatively few molecules that lie in the direction determined by θ and ϕ is

$$\mathrm{W} = \frac{1}{f(c_1) \ldots \ldots f(c_k)}, \quad . \quad . \quad . \quad . \quad (29)$$

and their total energy is

$$\mathrm{E}(\theta, \phi) = \tfrac{1}{2}m(c_1^2 + c_2^2 + \ldots \ldots + c_k^2) \quad . \quad . \quad (30)$$

Now multiply (30) by the undetermined constant λ and add to (29), and set the partial derivatives of this sum equal to zero to get the equilibrium state of maximum W or minimum E according to the Lagrange method :

$$-\mathrm{W}\frac{d}{dc_i}(\log f(c_i)) + \lambda m c_i = +\mathrm{W}\frac{d}{dc_i}(\log r_i) + \lambda m c_i = 0 \quad . \quad (31)$$

Here it is especially to be noted that the variations in the c's only occur in one direction, *i. e.*, θ and ϕ remain constant. Comparison of (26) and (31) shows that the latter is satisfied if

$$\lambda = \frac{-\mathrm{W}}{\tfrac{1}{2}m a^2}\left(\frac{1-p\cos^2\theta - q\sin^2\theta\cos^2(\phi \pm b)}{\tan^2 b}\right) = \frac{-\mathrm{W}}{k\mathrm{T}}\left(\frac{1-p\cos^2\theta - q\sin^2\theta\cos^2(\phi \pm b)}{\tan^2 b}\right). \quad . \quad (32)$$

$\tfrac{1}{2}m a^2$ appears here as an arbitrary constant and is replaced by kT, which is its value as shown by (61) below.

Using this value of λ, we may write eqs. (31) in the form

$$kT\frac{\tan^2 b}{1-p\cos^2\theta-q\sin^2\theta\cos^2(\phi\pm b)}$$

$$=\frac{mc_1dc_1}{dr_1/r_1}=\frac{mc_2dc_2}{dr_2/r_2}=\ldots\ldots\frac{mc_kdc_k}{dr_k/r_k} \quad . \quad . \quad . \quad . \quad (33)$$

Eq. (33), like (25), shows that there is a constant ratio, *i. e.*, an elastic modulus, between energy changes per molecule and corresponding strain in velocity space. This, of course, refers here to a particular direction determined by θ and ϕ, which were kept constant during the indicated variations on the right of (33). Evidently the left member, which is the elastic modulus, will vary for different directions. Hence we see that from the elasticity point of view the distribution is non-isotropic, that is, the elastic modulus varies with change in direction, though it is constant for each particular direction.

§11. *The Viscosity Distribution and the Equilibrium Equations between Stress and Strain in Velocity and Momentum Space, using Cartesian Coordinates.*

We have just shown, by using polar coordinates, that for each direction separately the steady state under consideration is determined by equilibrium equations between stress and strain. We now use cartesian coordinates (see (8)) and consider all directions together, and show that the viscosity distribution represents, like the Maxwell distribution, a state of equilibrium between stress and strain, being, moreover, the distribution of maximum entropy and strain for a given energy restricted, however, by the additional outside mechanical effect of the viscosity shear law (5) that, of course, does not disturb the Maxwell distribution. This is like the statics of indeterminate frames * where the principle of least work is always applied *subject* to the restrictions of elementary statics, which curtail the extent that the energy can be minimized.

* Kimball, *loc. cit.*

For the case in hand the energy before and after impact respectively is given by

$$\mathrm{E} = \Sigma\tfrac{1}{2}m(u^2+v^2+w^2) = \Sigma\tfrac{1}{2}m((u'+aw')^2+v'^2+w'^2)$$

$$= \mathrm{E}'+\mathrm{E}'' = \frac{nk\mathrm{T}}{2}(3+a^2),$$

where $\mathrm{E}' = \Sigma\tfrac{1}{2}m(u'^2+v'^2+w'^2) = \frac{3}{2}nk\mathrm{T},$

and

$$\mathrm{E}'' = \Sigma(\pm mau'w' + \tfrac{1}{2}ma^2w'^2) = \Sigma(\pm mauw - \tfrac{1}{2}ma^2w^2)$$

$$= \Sigma\tfrac{1}{2}ma^2w^2 = \Sigma\tfrac{1}{2}ma^2w'^2 = \frac{nk\mathrm{T}a^2}{2} \quad . \quad . \quad . \quad (34)$$

Here E′ is the energy of agitation that exists independent of viscosity since it has no constant a. It is the energy of agitation of the corresponding Maxwellian distribution of spherical symmetry. E″, on the other hand, is the additional energy impressed on the gas by the operation of the outside constant field of force according to the viscosity shear law (5). The right members show the summations carried out by integration according to eqs. (62), (73), and (75) below. In applying the principle of least internal work (or the equivalent principle of maximum entropy, probability, and strain) this E″ plays the role corresponding to the restrictions of elementary statics, *i. e.*, it curtails the extent to which the total energy, E, can be minimized.

We use the geometrical expression for weight like (23) for three dimensions :

$$\mathrm{W} = n^n(r_1r_2 \ldots r_n) = \frac{1}{(f(c_1)f(c_2)\ldots f(c_n))}, \quad . \quad . \quad (35)$$

and the Lagrange method for conditional maxima and minima, forming the function : $\mathrm{F} = \mathrm{W} + \lambda\mathrm{E} + \mu\mathrm{E}''$, where λ and μ are undetermined constants. In taking the variation of F in the X, Y, and Z directions we may keep in mind the fact that, in spite of its form, all the force represented by E″ is in the X direction, since $\pm aw$ is a correction term for the x component of velocity. Thus we have $3n$ equations for the three directions :

$$\left.\begin{aligned} \mathrm{W}\frac{dr}{r}+\lambda mu\,du-\mu m(a^2w\,dw\pm(au\,dw+aw\,du))=0,\\ \mathrm{W}\frac{dr}{r}+\lambda mu\,dv=0,\\ \mathrm{W}\frac{dr}{r}+\lambda mu\,dw=0. \end{aligned}\right\}. \quad (36)$$

If λ is $-\mathrm{W}/k\mathrm{T}$ and μ is $+\mathrm{W}/k\mathrm{T}$, these equations evidently give our distribution function (26) for velocities before and after impact using (61) below, which in cartesian form is

$$f(c)=\frac{1}{nr}=\mathrm{A}\exp\left[\frac{-m(u^2+v^2+w^2+a^2w^2\pm 2auw)}{2k\mathrm{T}}\right] \quad . \quad (37)$$

Accordingly the equilibrium equations (36) may be put in the form (replace λ and μ by the above values and transpose):

$$\left.\begin{aligned} k\mathrm{T}=\frac{m(u\pm aw)(du\pm a\,dw)}{dr/r}=\frac{mu'du'}{dr/r},\\ k\mathrm{T}=\frac{mv\,dv}{dr/r}=\frac{mv'dv'}{dr/r},\\ k\mathrm{T}=\frac{mw\,dw}{dr/r}=\frac{mw'dw'}{dr/r}. \end{aligned}\right\} \quad . \quad . \quad (38)$$

These are the stress-strain relations per molecule in velocity and momentum space at the free path ends (before and after impact) that determine our ellipsoidal viscosity distribution function (37) that applies before and after impact. This distribution function could also have been obtained by taking the partial derivatives of $\mathrm{F}=\mathrm{W}+\lambda\mathrm{E}+\mu\mathrm{E}''$ with respect to u, v, and w, and separately setting each equal to zero as they stand, instead of keeping the aw correction term's derivative in the first equation as in (36). This, however, would not be so illuminating in the physical sense as the above method, since aw stands for a velocity in the x direction. It is also evident from (38) that these stress-strain relations are the same as for a Maxwell distribution of spherical symmetry having velocity components u', v', and w', and that (38) could have been obtained by setting equal to zero the partial derivatives of a function $\mathrm{F}=\mathrm{W}+\lambda\mathrm{E}'$ with respect to u', v', and w'.

If we multiply both members of (38) by the number of molecules per c.c., we obtain the corresponding stress-strain equations :

$$\left.\begin{aligned} p=nk\mathrm{T}&=\frac{nmu\,du-nma(\pm(u\,dw+w\,du)-aw\,dw)}{dr/r},\\ &=\frac{nmu\,du-a^2nmw\,dw}{dr/r}=\frac{f_x-f_v}{dr/r},\\ p=nk\mathrm{T}&=\frac{nmv\,dv}{dr/r}=\frac{f_y}{dr/r},\\ p=nk\mathrm{T}&=\frac{nmw\,dw}{dr/r}=\frac{f_z}{dr/r}. \end{aligned}\right\} \quad (39)$$

The first of eqs. (39) simplifies as indicated, since the sum of the cross product terms $u'w'$ vanishes according to (34) and (75). The modulus of elasticity appearing here is the simple pressure formula $p=nmc^2/3$, and f_x, f_y and f_z are for the three directions, the corresponding energy differences upon impact for the n molecules per unit volume being the stresses that act in velocity or momentum space to maintain the distribution in a steady state as previously shown. The expression f_v is given by

$$f_v=a^2nmw\,dw=a^2f_z=a^2\sqrt{2\pi e}\;k\mathrm{T}.$$

$$f_x=(1+a^2)\sqrt{2\pi e}\;k\mathrm{T}=(1+a^2)f_y=(1+a^2)f_z. \quad . \quad . \quad (40)$$

It is evidently the stress acting in velocity and momentum space due to viscosity in the form corresponding to the viscosity shear law (5). Eq. (39) shows $f_v=a^2f_z$, and since f_z and f_y are each the same as the stress in the Maxwell distribution for that temperature, we have substituted this stress * in the right member as indicated. Furthermore, since f_v needs to be subtracted from f_x to give the isotropic situation given by (38) and (39), it follows that f_x exceeds $f_y=f_z$ by the amount f_v as indicated in the second of eqs. (40). Eqs. (38) refer to individual ranges (molecular separations throughout velocity space). Eqs. (39), on the other hand, are n times the mean value of (38), so that f_v is a total mean effect per *unit volume* acting only in *velocity* space. It is not to be confused with F′ of (2), which acts in ordinary space on unit *mass* having velocity components (u, v, w) and hence this position in velocity space. It is a striking

* Kimball, *loc. cit.* p. 620, eq. (25).

fact that the stresses in momentum space given by (40) are related to each other in the same way as the corresponding normal stresses in ordinary space given by (66), (73), and (74) below, where p_{xx} also exceeds p_{yy} and p_{zz} by the factor $1+a^2$. And in all cases the stresses are computed from energy changes involved in equivalent *impacts*.

It should be noted that the isotropic situation represented by (38) and (39) is not inconsistent with the anisotropic equilibrium equations (33). Thus, for $\theta = \pi/2$ and $\phi = \pm b$, the elastic modulus of (33) takes its maximum value $k\mathrm{T}/\tan^2 b$, so that corresponding energy differences produce the least strain. The reason they do not cause greater strain is shown by (38) and (39), *i. e.*, their effect is opposed by the stresses corresponding to f_v in (41) due to reacting viscous force. Thus the nonisotropic situation, so far as stresses are concerned, represented by (33) is consistent with the isotropic volume strain situation indicated in the present section by (38) and (39), and the symmetric entropy expressions in section 19, by virtue of counter-balancing viscous forces like f_v in (40), acting in velocity space.

§ 12. Theorem I.—*The necessary and sufficient Condition with Accuracy corresponding to the Gas Law for maintenance of Equilibrium between Stress and Strain in Velocity Space is the Normal Law for Velocity Distributions in each Direction, its two Constants being Arbitrary Functions of the Direction Angles.*

We note that (26) is the so-called normal law for unidirectional velocity distribution keeping θ and ϕ constant, and, furthermore, we would still have some particular type of the normal law provided the coefficient of $-c^2$ in the exponential were any other function whatever of θ and ϕ, and besides this, if the constant A of (26) were taken to be a variable function of θ and ϕ, thus varying with direction but independent of c. Now a glance at eqs. (31) and (32) shows that for *any one* of these possible normal laws the function A (θ, ϕ) would drop out from the logarithmic derivative of (31), and that an appropriate choice (corresponding to the exponential term) of λ as a function of θ and ϕ independent of c would *satisfy* (31), as well as its equivalent (33)

(the latter being modified for different choices of λ). In short, the equilibrium eq. (31) between stress and strain for any particular direction in velocity space gives on integration the normal law for velocity distribution in that direction with two "constants" of integration $A(\theta, \phi)$ and $\lambda(\theta, \phi)$, which are *arbitrary* functions of the direction angles. Hence the theorem as stated above.

The arbitrariness of the two constants of integration λ and A (and hence the elastic modulus of (33)) as functions of θ and ϕ, holds, of course, with reference to the equilibrium differential equations (31) and (33). Other restrictions will curtail their arbitrariness and adapt them to the requirements of any particular type of distribution like boundary conditions upon the solutions of ordinary differential equations. Thus (21) and the familiar Maxwell distribution of spherical symmetry are particular solutions of the differential equations (31) whose complete solution involves two arbitrary functions of the direction angles θ and ϕ. Refinements such as the Pauli exclusion principle and the new statistics can be derived from equilibrium equations * differing from (33) and (31) by the correction terms of the new statistics. Hence they are not included in the normal law for distributions and are outside the scope of the above Theorem I., which can very likely be generalized so as to include them.

§ 13. Theorem II.—*In case of a steady State independent of Time, the necessary and sufficient Conditions required of a Distribution Function to make it an exact Solution of Boltzmann's Continuity Equation are that it represents* (1) *a Constant Value of n, the Number of Molecules per unit Volume at each Position in Ordinary Space, and* (2) *Equilibrium of Forces everywhere in Momentum Space and Ordinary Space.*

Following Jeans †, we write Boltzmann's equation :

$$\frac{d(nf)}{dt} = -\left(X\frac{\partial}{\partial u} + Y\frac{\partial}{\partial v} + Z\frac{\partial}{\partial w} + u\frac{\partial}{\partial x} + v\frac{\partial}{\partial y} + w\frac{\partial}{\partial z}\right) nf + \left(\frac{d(nf)}{dt}\right)_1 . \quad (41)$$

* Kimball and Berry, Phil. Mag. xiii. p. 1131 (1932).

† 'Jeans, Dynamical Theory of Gases,' ch. viii. eq. (513).

We see, however, by (27) that nf is the reciprocal of the "range," and using (27) we may substitute r for the dependent variable nf:

$$\frac{dr}{dt}=-\Big(\mathrm{X}\frac{\partial}{\partial u}+\mathrm{Y}\frac{\partial}{\partial v}+\mathrm{Z}\frac{\partial}{\partial w}+u\frac{\partial}{\partial x}+v\frac{\partial}{\partial y}+w\frac{\partial}{\partial z}\Big)r+\Big(\frac{dr}{dt}\Big)_{\text{coll.}} \quad . \quad (42)$$

The factors $-1/r^2$ and $-e^{-s}$, which appear when (27) or (44) are used to replace nf in (41), are included implicitly in the collision term of (42) or (45) respectively. Boltzmann's equation is thus replaced by (42), its equivalent which represents time rate of change of the molecular spacing or range, r, at any place in phase space.

Again, we may take the product of all the ranges and get for the "Weight":

$$\mathrm{W}=n^n(r_1r_2\ldots r_n)=e^{\mathrm{Y}}=n^ne^{s_1+s_2+s_3\cdots s_n} \quad , \quad . \quad . \quad (43)$$

where Y is the total strain *, being the sum of the strains s_i (each weighted as indicated) which correspond to the various ranges:

$$s_i=\log r_i\,;\; r=\frac{1}{nf}=e^s\,;\; ds=dr/r. \quad . \quad . \quad . \quad (44)$$

These connexions—(43) and (44)—between strain and range show that (42), and hence Boltzmann's equation, is not only a continuity equation, but depends on the time rate of change of the *strain*. Its vital mechanical significance becomes more clear if we use (44) to replace the dependable variable in (41) or (42) by s:

$$\frac{ds}{dt}=-\Big(\mathrm{X}\frac{\partial}{\partial u}+\mathrm{Y}\frac{\partial}{\partial v}+\mathrm{Z}\frac{\partial}{\partial w}+u\frac{\partial}{\partial x}+v\frac{\partial}{\partial y}+w\frac{\partial}{\partial z}\Big)s+\Big(\frac{ds}{dt}\Big)_{\text{coll.}} \quad . \quad (45)$$

The s appearing in (45) is the strain appearing in the equilibrium equations (25), (31), (33), (38), being the logarithm of the range as indicated by (44). Thus Boltzmann's equation represents rate of change of strain as well as range, and also rate of change of entropy †.

* Kimball, *loc. cit.*
† Kimball, *loc. cit.* eq. (9).

We consider only the steady state where conditions do not vary with time, and the left member is zero and the right member includes the various effects which must sum up to zero. If n, the number of molecules per unit volume, is constant at each position throughout the gas (though different at different positions), then the molecular spacing * or range ($r=1/n$ for ordinary space) is a constant independent of time and $dr/dt=0$, and Boltzmann's equation is satisfied so far as ordinary space is concerned. This is the first requirement of Theorem II. for an exact solution of Boltzmann's equation. It is satisfied if the molecular flux is zero in all directions when referred to a suitable reference frame :

$$\int u\,dn=\int v\,dn=\int w\,dn=0. \quad . \quad . \quad . \quad . \quad (46)$$

As regards momentum space, we conclude that there will be no change of r, the range, so long as mechanical equilibrium prevails in momentum as well as ordinary space. This conclusion is supported by statistical mechanics as well as elasticity theory and statics, for it has been shown † how the Principle of Maximum Probability evoked by statistical mechanics is mathematically equivalent to Castigliano's Principle of Least Work used in statics, and that in both cases the corresponding equilibrium equations represent a balance between stress and strain according to elasticity theory. When Boltzmann's equation is replaced by the equivalent (42) or (45) it is clear that it must be satisfied by any distribution which provides complete mechanical equilibrium in momentum and in ordinary space with unchanging n the number of molecules per unit volume, because in this case we assume that *nothing* is *left* to cause a change in range or strain. Eqs. (25), (31), (33), and (38) show how the distribution of r's throughout momentum space is maintained by stresses, and we *assume* that we may thus *rely* on elasticity theory, and make the contention, tentatively, that nothing else but such agencies expressible as elastic effects can cause the changes represented by Boltzmann's equation.

* Kimball, "Entropy and Probability," Journ. of Phys. Chem. xxxiii. pp. 1558–1567 (1929).

† Kimball, *loc. cit.*

§ 14. Corollary.—*The Ellipsoidal Viscosity Distribution Function is an exact Solution of Boltzmann's Equation.*

We show in section 16 below that the mass flow or drift velocity is zero in every direction when referred to a reference frame moving with the gas. Thus (46) is equivalent to (56) below, and the number of molecules per c.c. is everywhere constant.

In section 18, the pressure in various directions is considered, and it is shown how equilibrium of forces is maintained in ordinary space. Eqs. (31), (33), and (36) show how the equilibrium of forces is maintained in velocity and momentum space. Since all the requirements of Theorem II. are thus satisfied, we have an exact solution of Boltzmann's equation.

To show this in more detail we take the ellipsoidal distribution function (21) applicable before or after impact, and let it be referred to a fixed reference frame. We have

$$f=\frac{1}{a^3\pi^{3/2}}\exp\left[-\frac{(u-u_0)^2+v^2+w^2+a^2w^2\pm 2a(u-u_0)w}{a^2}\right], \quad . \quad . \quad . \quad (47)$$

where $u_0=-Bz$ is the drift velocity of the gas being proportional to z. Now it is well known that the familiar Maxwellian or spherically symmetrical distribution causes the vanishing of the *collision term* of (41), whereas (47) does not, since it is ellipsoidal and not expressible * as

$$\log f=a_1(u^2+v^2+w^2)+a_2u+a_3v+a_4w+a_5.$$

We may, however, show in detail how (47) satisfies (41) by first writing (41) in the reduced form :

$$\begin{aligned} & w\frac{\partial(nf)}{\partial z} \\ & =\left(\frac{\partial(nf)}{dt}\right)_{\text{coll.}}=\left(\frac{\partial nf}{\partial u}\frac{du}{dt}+\frac{\partial nf}{\partial v}\frac{dv}{dt}+\frac{\partial nf}{dw}\frac{dw}{dt}\right)_{\text{coll.}} \quad (48) \\ & =\mathrm{X}'\frac{\partial nf}{\partial u}+\mathrm{Y}'\frac{\partial nf}{\partial v}+\mathrm{Z}'\frac{\partial nf}{\partial w}=\mathrm{X}'\frac{\partial nf}{\partial u}. \end{aligned}$$

* Jeans, *loc. cit.* eq. (24), chap. ii.

The terms involving X, Y, Z, the force field components *between* impacts, are all zero, since there is no such field, and the x and y partial derivatives also vanish since (47) depends only on z. The collision term depends on the internal viscous force components per unit mass at (u, v, w) in velocity space, which may be written X′, Y′, and Z′, and then by Newton's law we can replace the accelerations as indicated. X′ is evidently the reaction equal and opposite to the impressed internal force per unit mass referred to by (2) that is measured by the drift velocity gradient, whereas Y′ and Z′ are zero since there is no viscous force in these directions. Thus (41) may correctly be specialized for the case in hand to the extent of (48). In this form Boltzmann's equation says nothing about intermolecular forces or the nature of collisions, since the factor X′ given by (48) is merely Newton's law applied to large scale mass flow effects.

Now use (47) in (48) and get

$$\frac{2wf}{a^2}(u-u_0\pm aw)\frac{du_0}{dz}=-\mathrm{X}'\frac{2f}{a^2}(u-u_0\pm aw).$$

This simplifies to

$$\mathrm{X}'=-w\frac{du_0}{dz}=-\frac{du_0}{dt}=-\mathrm{F}'. \quad . \quad . \quad . \quad (49)$$

Hence this corollary shows how Boltzmann's equation is exactly satisfied and reduces to Newton's law given by (49), which is (2). It also illustrates in some detail the conclusions of Theorem II. as evidenced by (42) and (45), that the vanishing of Boltzmann's equation involves (for the steady states considered) (1) zero changes in entropy, strain and range at (u, v, w) in velocity and ordinary space, and (2) mechanical equilibrium to prevent any such changes. Furthermore, it appears that Boltzmann's equation says nothing in its own right about intermolecular forces, collisions, etc., since large scale Newtonian mechanics and elasticity theory can satisfy it exactly. It is here shown above to be satisfied by the distributions (47) *before* and *after* impact alike, since it is these distributions that are most immediately concerned with the collision term of Boltzmann's equation. It will readily be seen, however, that it will be satisfied by any one of the distributions (81) using any one of the a's given by (85).

§ 15. *Evaluation of the Constant* A.

Referring once more to the distribution (21) having the velocity ellipsoid in standard polar form, we proceed to evaluate the integral

$$dn = n\mathrm{A} \sin\theta d\theta d\phi \int_0^\infty \exp\left[\frac{-c^2(1-p\cos^2\theta - q\sin^2\theta\cos^2(\phi\pm b))}{a^2\tan^2 b}\right] c^2 dc \,. \quad (50)$$

keeping θ and ϕ constant. This gives

$$dn = \frac{n\mathrm{A}a^3\sqrt{\pi}\tan^3 b \sin\theta d\theta d\phi}{4(1-p\cos^2\theta - q\sin^2\theta\cos^2(\phi\pm b))^{3/2}}, \quad . \quad (51)$$

for the entire number of molecules dn included within the solid angle $\sin\theta d\theta d\phi$ before and after impact, taking into account all velocity magnitudes by virtue of the integration (50). Since the radical in the denominator of (51) is cubed, we see (having regard to (20)) that the left member dn is inversely proportional to the cubes of velocity magnitudes for a given solid angle. This checks the inverse cube relation pointed out in § 4. It was there shown to hold for any particular equal probability ellipsoid, whereas (51) merely shows that it also holds when all these particular ellipsoids are taken together as, of course, consistency requires that it must.

To integrate (51) over the angular ranges we first let ϕ = constant, and replace $\sin^2\theta$ by $1-\cos^2\theta$ and then write (51) in the form

$$dn = \frac{-n\mathrm{A}a^3\sqrt{\pi}\tan^3 b d\phi}{4}\int_1^{-1}\frac{du}{(\mathrm{P}-\mathrm{Q}u^2)^{3/2}};$$

where

$$u=\cos\theta,\ \mathrm{P}=1-q\cos^2(\phi\pm b),\ \mathrm{Q}=p-q\cos^2(\phi\pm b), \quad . \quad . \quad . \quad (52)$$

where the limits -1 and 1 for u correspond to the range π and 0 for the angle. This integral is easily evaluated by an elementary trigonometric substitution and gives

$$dn = \frac{n\mathrm{A}\sqrt{\pi}a^3\tan^3 b d\phi}{\sqrt{\mathrm{P}-\mathrm{Q}}2\mathrm{P}} = \frac{n\mathrm{A}\sqrt{\pi}a^3\tan^2 b d\phi}{2(1-q\cos^2(\phi\pm b))}, \quad . \quad (53)$$

where it will be noted that the radical $\sqrt{\mathrm{P}-\mathrm{Q}} = \tan b$ by (20) and (52), and divides into the numerator as

indicated. Finally (53) is to be integrated over the range 0 to 2π for the angle ϕ. Multiply numerator and denominator by $\sec^2(\phi \pm b)$ and use in the denominator the relation $1 + \tan^2 z = \sec^2 z$, then

$$n = \frac{n\mathrm{A}\sqrt{\pi}\,a^3 \tan^2 b}{2} \int_0^{2\pi} \frac{\sec^2(\phi \pm b)\, d\phi}{1 - q + \tan^2(\phi \pm b)}$$

$$= \frac{n\mathrm{A}\sqrt{\pi}\,a^3 \tan^2 b}{2\sqrt{1-q}} \arctan \left[\frac{\tan(\phi \pm b)}{\sqrt{1-q}} \right]_0^{2\pi} = n\mathrm{A}(a\sqrt{\pi})^3. \quad (54)$$

Here $\sqrt{1-q} = \tan^2 b$ by (20), and it will be noted that as ϕ ranges from 0 to 2π, the indicated arctan in the right of (54) traverses an equal interval. Comparison of the extreme right and left members of (54) evaluates the constant A :

$$\mathrm{A} = 1/a^3 \pi^{3/2} = \frac{1}{(a\sqrt{\pi})^3} \quad . \quad . \quad . \quad . \quad . \quad (55)$$

By using $\pm b$ in (50) and (51) we restrict consideration to the velocities of the n molecules per c.c. just before and after impacts take place. The same constant A, however, is given by letting b vary with x according to (85) and using the minus sign above. Then instead of (21) we would have

$$dn = n\mathrm{A} \exp \frac{-c^2(1 - p\cos^2\theta - q\sin^2\theta\cos^2(\phi - b)}{a^2 \tan^2 b} \times c^2 dc \sin\theta\, d\theta\, d\psi\, dx \quad . \quad (21\,a)$$

Here b, and hence p and q, are functions of x according to (85) and 20) (see (81) below also). Here (21 a) represents the number of molecules per unit volume, per element $d\omega = c^2\, dc \sin\theta\, d\theta\, d\phi$, and which, furthermore, have pursued the x fractional part of their free paths (the fraction between x and $x + dx$). It is a higher order infinitesimal than (21) and represents the distribution in velocity space at *any one time* of the $n\,dx$ molecules having traversed this x fractional part of their free paths. On the other hand (21), and hence (50) and (54), do not represent a distribution at any one time, since all n molecules per c.c. do not have impacts at the same time but (21) merely tells how the n molecules per c.c. will be distributed in velocity space whenever impacts do occur. If we use (21 a) in (50) instead of (21), the ensuing

integrations are all the same, although they now refer to the $n\,dx$ molecules which at a given instant have traversed the x fractional part of their free paths. And instead of (54) we get

$$dn = n\mathrm{A}(a\sqrt{\pi})^3\,dx$$

and
$$n = \int_0^1 n\mathrm{A}(a\sqrt{\pi})^3\,dx = n\mathrm{A}(a\sqrt{\pi})^3, \quad . \quad . \quad (54\,a)$$

since A and a are independent of x. Thus (54 a) gives the same value for A as (54) does.

§16. *The Mass Flow of Gas in each Direction is Zero when referred to the Reference Frame moving with the Gas.*

If the origin of coordinates and associated coordinate planes move along with the gas, it is obvious that the mass flow of gas with reference to such a system is *ipso facto* zero. This fact appears mathematically in that the strain ellipsoids, which are equal probability ellipsoids (20), are symmetric with respect to the origin, so that the probability of a velocity magnitude in any particular direction is equal and opposite to such a probability in the exact reverse direction. A mathematical check can be made by calculating the mass flow in any direction. Along the Y-axis, we have $c\cos\theta$ for the component of each velocity in that direction. The corresponding flux will then be given by

$$\int dn\, c\cos\theta$$

$$= \int n\mathrm{A}\exp\left[\frac{-c^2(1-p\cos^2\theta - q\sin^2\theta\cos^2(\phi\pm b)}{a^2\tan^2 b}\right] \times c^3\sin\theta\cos\theta\,d\theta\,d\phi\,;$$

$$= \frac{-n\mathrm{A}a^4\tan^4 b}{8}\,d\phi\int_0^{\pi}\frac{d(\cos 2\theta)}{(1-p\cos^2\theta - q\sin^2\theta\cos^2(\phi\pm b))^2}\,;$$

$$= \mathrm{F}(\cos 2\theta)_0^{\pi} = 0 \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (56)$$

Here the denominator of the middle expression (56) can be written $(\mathrm{A}' - \mathrm{B}'\cos 2\theta)^2$, since $\sin^2\theta$ and $\cos^2\theta$ are each expressible linearly in terms of $\cos 2\theta$. Hence the integral will be a function of $\cos 2\theta$ only (disregarding ϕ), which reduces to zero when the upper and lower limits

π and o are introduced with their equal values of $\cos 2\theta$. Likewise, if we use (21) to calculate the mass molecular flux in the x or z directions. Thus $\int dn\, c \sin\theta \cos\phi = 0$, and $\int dn\, c \sin\theta \sin\phi = 0$. These latter integrations are not so easy as for the y component of velocity, and it is necessary to integrate with respect to ϕ as well as θ, using substitutions like that of the last and subsequent sections. In both cases, however, the definite integral will be seen to vanish because the function emerging has the same value at upper and lower limits. These results apply evidently to the distribution (81) between impacts as well as to (21) that refers to this situation before and after impact.

§ 17. *Temperature and the Viscosity Distribution Function.*

Since the components of agitation velocity in the y and z directions have no viscous force to disturb them, the associated mean energy of agitation will for these two directions be $k\mathrm{T}/2$, as it is for the various degrees of freedom in the Maxwellian or spherically symmetrical distribution. Thus for the energy components in the y direction we have

$$\int dn \frac{mc^2}{2} \cos^2\theta = nk\mathrm{T}/2 \quad . \quad . \quad . \quad . \quad . \quad (57)$$

to determine the as yet unknown constant a. For brevity, we drop the $m/2$ and consider first the integral

$$\int dn\, c^2 \cos^2\theta = \iiint_0^\infty n\mathrm{A} \exp \times \left[\frac{-c^2(1-p\cos^2\theta - q\sin^2\theta\cos^2(\phi \pm b))}{a^2 \tan^2 b}\right] c^4 dc \sin\theta \cos^2\theta\, d\theta\, d\phi \quad . \quad (58)$$

$$= \int_0^{2\pi} \frac{3n a^2 \tan^5 b\, d\phi}{8\pi} \int_{-1}^{+1} \frac{u^2 du}{(\mathrm{P}-\mathrm{Q}u^2)^{5/2}} = \frac{n a^2 \tan^2}{4\pi} \times \int_0^{2\pi} \frac{d\phi}{1-q\cos^2(\phi \pm b)} . \quad . \quad (59)$$

The left member comes from integrating (58) with respect to c and using the substitutions (55) and (52) for A, u,

P, and Q. The right member results from integrating the left with the help of $Qu^2 = P \sin^2 z$, and using (52) again. The *integral* on the right is the same as that in (53) and is $2\pi/\tan^2 b$, which when substituted gives

$$\int dn\, c^2 \cos^2 \theta = n\alpha^2/2 \quad . \quad . \quad . \quad . \quad . \quad (60)$$

Comparison of (57) and (60) shows that

$$\tfrac{1}{2} m\alpha^2 = k\mathrm{T}, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (61)$$

which shows that the α here used is the most probable velocity of the corresponding Maxwell distribution at temperature T.

Likewise, for the energy component in the z direction, we get (61) for the relation between α and T, by evaluating (compare with the next section):

$$\int dn \frac{mc^2}{2} \sin^2 \theta \sin^2 \phi = \frac{nk\mathrm{T}}{2} = \frac{1}{4} nm\alpha^2 \quad . \quad . \quad (62)$$

As before, the integration is best carried out first with respect to c, then θ, and finally ϕ, using the same substitutions as above for θ. The last integration requires the expansion of $\sin^2 \phi$:

$$\sin \phi = t \sin(\phi \pm b) \pm s \cos(\phi \pm b)\,; \quad \text{where} \quad \begin{matrix} s = \sin b \\ t = \cos b \end{matrix} \quad (63)$$

And the cross product term that comes from the square of (63) yields a function of $\cos 2(\phi \pm b)$ which vanishes like (56) when the limits are substituted and the squared terms are taken care of by multiplying numerator and denominator by $\sec^4(\phi \pm b)$, and using the transformation

$$\tan(\phi \pm b) = \sqrt{1-q}\, \tan x \quad . \quad . \quad . \quad . \quad (64)$$

and noting that as ϕ ranges from 0 to 2π, x traverses an equal interval between its limits. Thus we see by (61) that the energy of these two components for the distribution (21) is related to temperature and the constant α in the same way as the energy components and the constant α of the Maxwell distribution are related to temperature.

§ 18. *Pressure and the Viscosity Distribution Function.*

It is well known that in case of the Maxwellian distribution of velocities the pressure is given by the integral

of twice the energy of the molecules in one translational degree of freedom. Thus :

$$p=\int dn\, mu^2=\int dn\, mc^2\cos^2\theta=nm\alpha^2/2=nk\mathrm{T}, \quad . \quad (65)$$

where the middle integral shows an integration of twice the energy components along the polar axis of coordinates. This also can be viewed as the flux of momentum or as the integral of the product of momentum change $2mu$ by impact frequency $\frac{u}{2}$. In case of the viscosity distribution we obtain the pressure formulæ in the same way. Thus for a pressure in a direction at right angles to the direction of viscous drag we have only to multiply (57) and (62) by 2 to get formulæ corresponding exactly to (65) for the pressure in the y and z directions, which is unaffected by the viscous force

$$p_{yy}=p_{zz}=\int dn\, mc^2\cos^2\theta=\int dn\, mc^2\sin^2\theta\sin^2\phi=nk\mathrm{T}. \quad . \quad . \quad . \quad . \quad (66)$$

The pressure in the direction that the viscous force acts will obviously be greater than (66) due to extra energy imparted to the molecules by the viscous force. To calculate it we have to evaluate the integral like (66) corresponding to the x direction :

$$p = \int dn\, mc^2\sin^2\theta\cos^2\phi=nm\mathrm{A}\iiint \exp \times\left[\frac{-c^2(1-p\cos^2\theta-q\sin^2\theta\cos^2(\phi\pm b))}{\alpha^2\tan^2 b}\right] \times c^4\sin^3\theta\cos^2\phi\, dc\, d\theta\, d\phi \quad . \quad (67)$$

The integrations here with respect to c and θ are the same as (62), but we now have $\cos^2\phi$ instead of $\sin^2\phi$ in the last integrand. Thus :

$$p_{xx}=\frac{nm\alpha^2\tan^4 b}{2\pi}\int_0^{2\pi}\frac{\cos^2\phi\, d\phi}{(1-q\cos^2(\phi\pm b))^2} \quad . \quad (68)$$

Instead of (63), we now have to use the substitution for $\cos\phi$:

$$\cos\phi = t\cos(\phi\pm b)\pm s\sin(\phi\pm b);\ \text{where}\ \begin{matrix} s=\sin b \\ t=\cos b \end{matrix}, \quad (69)$$

and when (69) is squared and substituted in (68), the

cross product term again yields a vanishing integral like (56), and the squared terms give

$$p_{xx}=-\frac{nm\alpha^2\tan^4 b}{2\pi}\int_0^{2\pi}\frac{t^2\cos^2(\phi\pm b)+s^2\sin^2(\phi\pm b)d\phi}{(1-q\cos^2(\phi\pm b))^2}. \quad (70)$$

Multiply numerator and denominator by $\sec^4(\phi\pm b)$, and then under the transformation (64) as before, we have

$$p_{xx}=\frac{nm\alpha^2\tan^4 b}{2\pi(1-q)^{3/2}}\int_{x_0}^{x_1}(t^2\cos^2 x\,dx+s^2(1-q)\sin^2 x)dx. \quad (71)$$

For the limits in question the definite integrals of $\sin^2 x$ and $\cos^2 x$ are each equal to π as in case of (62), so that, taking account of (20) and (69) to replace q, s, and t, we have

$$p_{xx}=\frac{nm\alpha^2}{2}\left(\frac{\cos^2 b+\sin^2 b\tan^4 b}{\tan^2 b}\right), \quad (72)$$

and the trigonometric factor is seen by elementary (though involved) trigonometry to be identically equal to $1+(\cot b-\tan b)^2=1+a^2$. Thus

$$p_{xx}=\frac{nm\alpha^2}{2}(1+a^2)=nk\mathrm{T}(1+a^2)=p(1+a^2) \quad (73)$$

shows how the pressure in the x direction is increased in terms of the proportionality factor a being the amount of shear given by (16). The difference between p_{xx} and $p=nk\mathrm{T}$ is evidently the measure of the increase in normal pressure due to viscous drag, and it is of a strikingly simple form :

$$p_{\prime\prime}=p_{xx}-p=nk\mathrm{T}a^2=pa^2 \quad (74)$$

The integrations of this and the last sections may be checked by a calculation of the total energy which must equal (66) plus one-half of (73). Thus we find

$$\int dn\tfrac{1}{2}mc^2=\frac{nm\mathrm{A}}{2}\left\{\exp\left[\frac{-c^2(1-p\cos^2\theta-q\sin^2\theta\cos^2(\phi\pm b))}{a^2\tan^2 b}\right]\times c^4\sin\theta\,dc\,d\theta\,d\phi\right. \quad (75)$$

$$=\frac{nm\alpha^2}{4}(3+a^2)=\frac{nk\mathrm{T}}{2}(3+a^2);$$

where $nk\mathrm{T}a^2/2$ is evidently the excess energy of agitation

in the x direction maintained by the viscous force throughout the gas.

§ 19. *Entropy and the Viscosity Distribution Function.*

We have seen that the weight per molecule for our distribution function is given by (26), and hence the total weight is :

$$W = \frac{1}{f(c_1)f(c_2)\ldots.f(c_n)}, \quad . \quad . \quad . \quad . \quad . \quad (76)$$

being the geometrical expression for weight expressed in terms of the distribution function (21) :

$$f(c) = A \exp \frac{-c'^2}{a^2} = A \exp \left[\frac{-c^2(1-p\cos^2\theta - q\sin^2\theta\cos^2(\phi \pm b))}{a^2 \tan^2 b}\right] = \frac{\exp}{(a\sqrt{\pi})^3}\left[\frac{-mc'^2}{2k'T}\right] \quad . \quad (77)$$

using (55) and (61) in the right-hand member. If we consider momentum space instead of velocity space, an m^3 must be included in the volume element $d\omega$ on the right of (21), and a corresponding one will appear in the denominator of (55). Hence, in view of (61), we have

$$A = \frac{1}{(ma\sqrt{\pi})^3} = \frac{1}{(2\pi mkT)^{3/2}} . \quad . \quad . \quad . \quad . \quad (78)$$

Using this value of A in (77) and substituting (77) in (76) gives

$$W = (2\pi mkT)^{\frac{3n}{2}} \exp\left[\frac{m(c'^2_1 + c_2'^2 + \ldots . c_n'^2)}{2kT}\right] . \quad (79)$$

for the weight of the n molecules under consideration. The exponent of e in (79) is evidently not the total energy of agitation divided by kT, but E'/kT where $E' = \Sigma\frac{1}{2}mc'^2$ is the total energy of the Maxwellian velocity distribution that corresponds under transformation (7) to the ellipsoidal viscosity distribution. The entropy is obtained by substitution of (79) in Boltzmann's equation

$$S = k \log W + C = R \log (2\pi mkT)^{3/2} + \frac{E'}{T} + C \quad . \quad (80)$$

It is striking that eqs. (79) and (80) are exactly the same as for the corresponding Maxwell distribution of spherical symmetry for this temperature.

To interpret this we recall that, according to section 5 above, the ellipsoidal distribution functions involve simple shear in velocity space with distortion, but no expansion, and, furthermore, the average shear is zero, so that the equivalence of (79) and (80) and the corresponding expressions for the Maxwell spherical distribution would seem to be justified. Thus for the distribution of velocities immediately *before* impact the "amount of shear" is $+a$, and likewise it is $-a$ for the distribution immediately *after* impact according to the fundamental shear law (5). One distribution is a mathematical reflexion or image of the other, and the sum of the shear effects is zero at *any one time*, since we do not get all the *after* effects at one time and all the effects just *previous* to impact at another time. And the fact that one is a mathematical reflexion of the other is merely Newton's third law of motion incorporated into the present theory.

§ 20. *The Viscosity Distribution includes, besides the Ellipsoidal ones before and after Impact, a Single Infinity of Distributions for Points along the Molecular Free Paths including the Spherical Maxwell Distribution for the Mid-Points between the Impacts.*

Consider, now, the velocity distributions *between* impacts, where mechanical effects are lacking, and note that the distribution before impact with $+a$ for its amount of shear arises out of the pure kinematics of the situation as represented by drift velocity and drift velocity gradient together with free paths, so that these kinematical factors operate to exactly reverse the correction or shear *impressed* on the molecules by viscosity as revealed in the distribution immediately *after* impact with its negative amount of shear $-a$. And it is by virtue of the free paths of moving molecules and consequent shifts of position that the distribution is transformed kinematically from one of negative shear $-a$ to one of positive shear $+a$. Furthermore, we note that symmetry requires for the half-way point in the free paths, that the impressed shear $-a$ shall be reduced to zero, and then we have the Maxwell distribution of spherical symmetry.

Hence the surprising result that our viscosity distributions actually include the Maxwell distribution of spherical symmetry, it being the one out of these distributions giving velocities in the *middle* of the free paths. Generalizing this idea, we find that the amount of shear of a distribution is *proportional* to the distance from the centre of the free path, being positive after the centre is passed and negative before it is reached, and being $+a$ at the end of the free trip (just before impact). For if pure kinematic effects due to *constant* speeds between impacts cause a half change of shear from $-a$ to 0 during the first half of the molecular free paths, and an entire change from $-a$ to $+a$ during the entire free path lengths, we seem obliged to conclude that this *uniform* kinematic effect causes changes in shear *proportional* to x, the fraction of the several molecular free paths traversed. It is a striking fact that viscosity represents a Maxwell distribution in the middle of its free paths, as well as two other distributions before and after impacts, which are mathematical reflexions or images of each other and obtained from the included Maxwell distribution by a simple shear $\pm a$ in velocity space. We may express these ideas mathematically by replacing "a" used in (8) or (10) by a variable "amount of shear," $a(2x-1)$, where x is the fractional part of the free path already traversed by the molecules of the particular distribution considered and ranges from zero to one. Then the distribution given by using the cartesian expression (8) in (77) is

$$f=\frac{1}{\pi^{3/2}a^3}\exp\left[\frac{-(u^2+v^2+w^2+a^2w^2(2x-1)^2-2auw(2x-1)}{a^2}\right]. \quad (81)$$

We see that, when $x=\frac{1}{2}$, (81) reduces to the Maxwell distribution that applies to the molecules at mid-points of the free paths. And when x is 1 or 0, we get the distributions here treated before and after impact respectively. We may still speak of (81) as a distribution, *i. e.*, *the* viscosity distribution function, but it is to be borne in mind that it includes a single infinity of ellipsoidal distributions including the Maxwell distribution and the two here specially treated.

§ 21. *The Mean Velocity Magnitudes.*

The magnitudes of the mean velocity components in each direction may be obtained in the usual way. For the y direction choose $\pi/2$ and 0 for the limits in (56) instead of the π and 0 shown there. These mean velocity magnitudes for components in the y and z directions turn out to be of the same form as the corresponding quantities in the Maxwell distribution of velocities. Thus, using (21),

$$\int_0^{2\pi}\int_0^{\frac{\pi}{2}}\int_0^{\infty} dnc\cos\theta = \frac{n}{2}\frac{a}{\sqrt{\pi}}; \quad \text{and} \quad \int_0^{\pi}\int_0^{\pi}\int_0^{\infty} dn\, c\sin\theta\sin\phi = \frac{n\,a}{2\sqrt{\pi}}. \quad . \quad (82)$$

and, there being $n/2$ molecules, the mean velocity components are

$$v = \bar{w} = a/\sqrt{\pi}. \quad . \quad . \quad . \quad . \quad . \quad . \quad (83)$$

The mean x component of velocity magnitude before or after impact, being increased by the viscous force, will, of course, be greater than (83). Thus

$$\int_{-\frac{\pi}{2}}^{\frac{\pi}{2}}\int_0^{\pi}\int_0^{\infty} dn\, c\sin\theta\cos\phi = \frac{n\,a}{2\sqrt{\pi}}\sqrt{1+a^2}\,; \quad \bar{u} = \frac{a}{\sqrt{\pi}}\sqrt{1+a^2}, \quad . \quad . \quad (84)$$

and

$$c = \frac{1}{n}\int_0^{2\pi}\int_0^{\pi}\int_0^{\infty} c\,dn = \frac{2a}{\sqrt{\pi}}\left(1+\frac{2a^2}{15}+a^4(\quad)+\ldots.\right)$$

The integrations here are carried out first with respect to c, then θ, and finally ϕ, using (16), (21), and (55). In case of $\bar{c}$ the last integration becomes an elliptic integral, expressible as the series

$$\bar{c} = \frac{a}{\sqrt{\pi}}\left\{1+\frac{\tan^4 b}{2\pi}\ldots.\right. \times \left.\int_0^{2\pi}\frac{d\phi}{P^2}\left[1+\frac{1}{3}\frac{Q}{P}+\frac{1}{5}\left(\frac{Q}{P}\right)^2+\frac{1}{7}\left(\frac{Q}{P}\right)^3+\frac{1}{9}\left(\frac{Q}{P}\right)^4+\ldots\right.\right.$$

where P and Q are given by (52). The integral is expressible as a power series in a^2, and terms of lower power

than a^4 are completely given by the first three terms of the series in the integrand (since by (16) and (20) Q is proportional to a). Since by (96), "a" is usually less than 10^{-6}, we see by (84) that c differs from $c' = \frac{2a}{\sqrt{\pi}}$ of the corresponding Maxwell distribution by less than 1 part in a million million.

The mean velocities given by (83) and (84) refer to velocities immediately *before* and *after* impact, since these distribution functions are used in the integrals. More important than $\bar{c}$ given by (84), however, is the general mean of all velocity magnitudes involved in (81) which will be less than (84) by virtue of the linear variation of the amount of shear between the values $-a$ and $+a$ according to the last section. Thus, recalling (16), "a"$=a(2x-1) = \cot b - \tan b$; and

$$\bar{a^2} = \int_0^1 a^2(2x-1)^2 dx = \frac{a^2}{3} \quad . \quad (85)$$

$$\bar{c} = \frac{2a}{\sqrt{\pi}}\left(1 + \frac{2a^2}{45} + (\quad)a^4\right).$$

This gives the true mean velocity magnitude for the viscosity distribution, (81) being the average value of the c's given by (84) when "a" varies linearly from the extremes of $-a$ to $+a$. The assumption underlying (85) is that there is equal probability for all the amounts of shear lying between $-a$ and $+a$. In view of the minuteness of "a" the $\bar{c}$ given by (85) also differs by less than 1 part in a million million from $c' = \frac{2a}{\sqrt{\pi}}$ of the included Maxwell distribution.

§22. *The Drift Velocity Gradient and Numerical Value of "a."*

To obtain the drift velocity gradient, we note that according to the viscosity shear law (5) positive z components of velocity, w, are associated *before* impact with a positive increment aw added to the x velocity component, and that negative z components give negative aw's, and that *after* impact the increment aw is of *opposite* sign from the z component w. Thus we obtain the mean forward velocity (in the x direction) just *before* impact or the backward velocity *after* impact (according as we

use $\phi-b$ or $\phi+b$) for molecules coming from *below* the xy plane by confining the first integral (84) to positive w's, *i. e.*, take 0 and π for the limits of ϕ. Thus carrying out the integrations for c and θ, we have

$$\int_0^{\pi}\int_0^{\pi}\int_0^{\infty} dn\, c \sin\theta\cos\phi = \frac{n\alpha\tan^3 b}{4\sqrt{\pi}}\int_0^{\pi}\frac{\cos\phi\, d\phi}{(1-q\cos^2(\phi\pm b))^{3/2}}.$$

When $\cos\phi$ is replaced by (69), this becomes two integrals, which integrate under the transformations

$$\left.\begin{aligned}\sqrt{1-q}\tan x &= \sqrt{q}\sin(\phi\pm b)\\ \sin y &= \sqrt{q}\cos(\phi\pm b)\end{aligned}\right\} \quad . \quad . \quad (86)$$

to

$$\frac{n\alpha\tan^3 b}{4\sqrt{\pi}}\left[\frac{t\sin x}{\sqrt{q(1-q)}} \pm \frac{s\tan y}{\sqrt{q}}\right]_{\phi=0}^{\varphi=\pi} \quad . \quad . \quad (87)$$

When the limits $\pi/2$ and $-\pi/2$ are used in (86) and (87) the latter reduces to (84). Using π and 0, however, and taking account of the a and the q defined by (16) and (20), gives

$$\int_0^{\pi}\int_0^{\pi}\int_0^{\infty} dnc\sin\theta\cos\phi = \pm\frac{n}{2}\frac{a\alpha}{\sqrt{\pi}} = \pm\frac{n}{2}a\bar{w}\,. \quad (88)$$

(When we use $\phi+b$ in (86) instead of $\phi-b$, the values of $\tan x$ and $\sin x$ for $\phi=\pi$ and $\phi=0$ both change signs, whereas $\sin y$ and $\tan y$ remain unaffected. Hence (87) and (88) are completely reversed in sign.) Since $\frac{n}{2}$ is the number of molecules coming from below the xy plane, their mean forward or backward velocity *before* or *after* impact respectively in the x direction is given by $\frac{a\alpha}{\sqrt{\pi}} = a\bar{w}$, being the product of the amount of shear (16) by the mean vertical velocity component (83). Likewise, using limits $\phi=2\pi$ and $\phi=\pi$ in (86) and (87) gives the negative of (88) for the mean backward or forward velocity before or after impact respectively for molecules coming *down* from *above* the xy plane. (See fig. 8 for graphical analysis of this situation.) The difference is

$$\Delta u_0 = \frac{2a\alpha}{\sqrt{\pi}} = -\lambda\frac{du_0}{dz} = 2a\bar{w}, \quad . \quad . \quad . \quad (89)$$

where λ is the mean free path. This equation of course defines the mean free path as the vertical distance that

separates layers of gas whose differences in drift velocity is Δu_0, this being taken as the difference between the mean backward velocity *after* impact of molecules coming from below, and the mean forward velocity *after* impact of molecules coming from above through the xy plane. It is a striking fact that it is the mean velocities *after* impact that determine the mass flow and drift velocity gradient, since they operate through the space intervals of their free paths during a corresponding time, whereas velocities *before* impact are expunged immediately on coming into existence. From (89) we have

$$a=2\cot 2b=-\frac{\sqrt{\pi}\lambda}{2a}\frac{du_0}{dz}=-\frac{\lambda}{c}\frac{du_0}{dz}=-10^{-10}\frac{du_0}{dz}\quad,\quad.\quad(90)$$

which checks (1) and shows again how the amount of shear a is $\frac{du_0}{dz}$ divided by the number of impacts per second, using $c=\frac{2a}{\sqrt{\pi}}$, which we saw differs from (85) by less than one part in a million million. The numerical term gives the order of magnitude of the constants a and b for a given velocity gradient. (See (96) below.) Thus for $\frac{du_0}{dz}=-1000$, a large velocity gradient, $a=10^{-7}$ and $\cot 2b=\frac{1}{2}10^{-7}$ which leads to $\pi/4-b=\cdot 005$ sec. of arc. This latter angle measures the rotation of the axes of the ellipsoid of the distribution functions from the spherically symmetric position which is theirs when $b=\pi/4$, and hence $a=0$. (See fig. 7.) Since $\frac{\pi}{4}-b$ is so minute, even for large velocity gradients, we see how very small is the departure from the Maxwellian distribution of spherical symmetry, and the two eccentricities e_1 and e_2 given by (20) are less than 10^{-3} for this a, e_2 being, as always, about $\sqrt{2}$ times e_1.

§ 23. *The Viscosity Coefficient and Elastic Modulus p connecting Shearing Stress and Amount of Shear.*

We may calculate the excess flux of momentum across unit area in the xy plane that gives rise to the internal tangential stress in the x direction, which is the viscous drag. $\int w\,dn$ is the numerical *molecular* flux across

unit area and $m\int w^2\,dn = \int m\,dn\,c^2\sin^2\theta\sin^2\phi = nkT = p_{zz} = p$ is the flux of mw across unit area, being the normal stress in the z direction according to (66). But maw is that part of the molecular momentum in the x direction due to viscosity. The amount of this crossing unit area in unit time for all molecules is, by comparison with (66),

$$ma\int w^2\,dn = pa = \frac{nma\,a^2}{2} = p_{xz} = -\mathrm{F}, \quad . \quad . \quad (91)$$

where p_{xz} represents tangential stress in the x direction on unit area normal to the z axis. It is a striking fact that (91) gives p, the pressure, as the proportionality factor connecting the tangential stress p_{xz} and the amount of shear "a." Thus the pressure p is the elastic modulus for the viscosity shearing stress and strain given by (91), as it is in (39), being the familiar bulk modulus for volume changes in ordinary space of gases obeying the gas law, as well as the elastic modulus connecting stress and strain in velocity space for both the Maxwell * and Fermi-Dirac † Distributions. Comparison of (90) and (91) shows that

$$\mathrm{F} = -p_{xz} = +\frac{\sqrt{\pi}}{4}n\lambda m\,a\frac{du_0}{dz} = +\frac{\pi}{8}n\lambda m\bar{c}\frac{du_0}{dz} = \eta\frac{du_0}{dz}, \quad (92)$$

where we have set $\bar{c} = \bar{c}' = \dfrac{2a}{\sqrt{\pi}}$, the mean velocity magnitude of the corresponding Maxwell distribution from which it differs by a negligible amount according to (85) above. F is the impressed force per unit area that causes the viscous flow. Eq. (92) shows that the viscosity coefficient is

$$\eta = \frac{\pi}{8}n\lambda mc \quad . \quad . \quad . \quad . \quad . \quad . \quad (93)$$

If we assume that this $\lambda = \dfrac{1}{\sqrt{2}\pi n\sigma^2}$ (Maxwell's mean free path), then the numerical factor $\pi/8$ involved is larger than the 1/3 given by the simplest viscosity theory, and smaller than ·499 given by Chapman using successive approximations ‡. This shows that the λ involved in

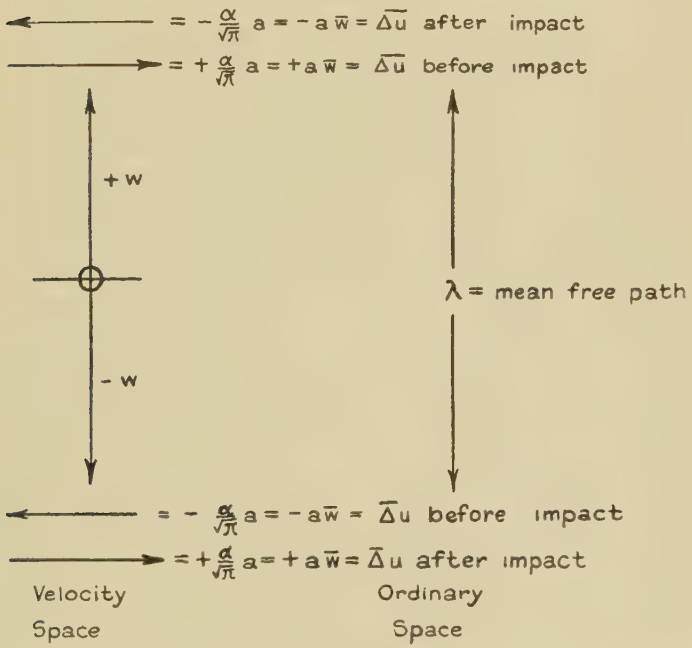
* Kimball, *loc. cit.* p. 619, eq. (24).
† Kimball and Berry, Phil. Mag. xiii. p. 1131 (1932), eq. (29).
‡ Jeans, 'Dynamical Theory of Gases,' chap. xi. eq. (773).

(89) approximates Maxwell's mean free path, and makes (93) an acceptable expression for the viscosity coefficient for ideal gases.

§ 24. *The Viscosity Couple.*

The ideas of the last two sections may be shown graphically as in fig. 8, where the mean velocity increments due to viscosity in the x direction given by (88)

Fig. 8.



The viscosity couple.

are shown both *before* and *after* impact for molecules coming from above $(-w)$ and from below $(+w)$. The equality of magnitudes before and after impact is required by Newton's third law of motion. Note that a *left*-hand twist in velocity space is indicated by the *after* impact strain ellipsoids graphed in figs. 6 and 7 above. And likewise an equal *right*-hand twist for the vectors *before* impact. Note also that it is the mean velocities *after* impact that are in the direction of drift above and below the xy plane and determine the gradient of drift as indicated

in section 22, since they obtain during the ensuing space and time interval of the free path, whereas velocities before impact are expunged immediately on arrival.

Since molecules crossing the xy plane with a *positive* w have next impacts above this plane and *vice versa*, we see that the momentum changes implied by the indicated velocity increments lost and acquired on impact shown in fig. 8 involve a viscous torque or right-hand twist which balances the impressed torque that causes the drift velocity and drift velocity gradient. For the upper force to the right we restrict (91) to positive w's, *i. e.*, take the limits of ϕ as π and 0, and get one-half of that integral : $pa/2$. Likewise, using 2π and π for the limits of ϕ in (91) we get $pa/2$ for the force to the *left* below the xy plane due to molecules crossing unit area from above. Since these forces are equal and opposite, they constitute a mechanical couple *. Now we have already defined the mean free path λ used here in section 22 as the vertical distance that separates layers of gas having differences of drift velocity $\Delta u_0 = \frac{2a\alpha}{\sqrt{\pi}}$ as determined by mean velocities *after* impact. It will thus be the same vertical interval shown in fig. 8 that separates the upper place where the equivalent mean excess positive change of momentum due to viscosity would occur, from the lower place where the equal mean excess minus change in momentum would occur in order to give the actual integrated effect. Hence the viscosity couple for the n molecules per unit volume is

$$\text{Couple}=\text{Torque}=2\left(\frac{pa\lambda}{2}\right)=\lambda p_{xz}=pa\lambda=\frac{n\lambda m a\alpha^2}{2}=n\lambda a k\mathrm{T} \quad . \quad . \quad . \quad . \quad (94)$$

which is proportional ($n\lambda$=const.) to "a" the amount of shear and to the absolute temperature. This couple doubtless plays an important role in the well-known and but partly understood phenomenon of turbulence.

§ 25. *The Viscosity Stress Matrix.*

It is well known that in case of a perfect fluid at rest, we have equal normal stresses and no tangential stresses :

* Kimball and King, Phil. Mag. xiii. p. 888 (1932), Theorem ii. p. 894.

$$\begin{Vmatrix} +p & 0 & 0 \\ 0 & +p & 0 \\ 0 & 0 & +p \end{Vmatrix}.$$

The corresponding stress matrix for the gas under consideration is according to (66), (73), and (91) above:

$$\begin{Vmatrix} +p(1+a^2) & 0 & +pa \\ 0 & +p & 0 \\ 0 & 0 & +p \end{Vmatrix} = \begin{Vmatrix} p_{xx} & 0 & p_{xz} \\ 0 & p_{yy} & 0 \\ 0 & 0 & p_{zz} \end{Vmatrix}. \quad (95)$$

Besides the familiar tangential stress $pa=p_{xz}=-\eta\frac{du_0}{dz}$ acting in the x direction on the face normal to the z direction, it is noteworthy that there is an increase in *normal* pressure in the x direction. The above tangential stress is positive—$\left(\frac{du_0}{dz}\right)$ is minus—and is balanced by $F=-p_{xz}$, the tangential impressed force passed down from above which drags the gas backward. The resultant tangential force is thus zero due to action and reaction, and the Cauchy symmetry of tangential stresses is maintained.

The increase in normal pressure pa^2 in the x direction sustained by the outside force is observed to be a second-order effect. We may compare its magnitude with that of the tangential stress $pa=-\eta\frac{du_0}{dz}$. Using c.g.s. units, we take for, say, helium gas at approximately normal conditions :

$$\left.\begin{aligned} p&=10^6 \text{ dynes/cm.}^2 \\ \eta&=2\times10^{-4} \\ -\frac{du_0}{dz}&=1000=10 \text{ metres/sec./cm.} \\ p_{xz}=pa&=-\eta\frac{du_0}{dz}=\cdot2 \text{ dynes/cm.}^2 \\ a&=2\times10^{-7} \\ \Delta p_{xx}=pa^2&=4\times10^{-8} \text{ dynes/cm.}^2 \end{aligned}\right\} \quad . \quad . \quad (96)$$

A velocity gradient of 1000/sec./cm. is very large. We conclude that ordinarily the amount of shear a is very minute (less than one millionth), and that the increase in the x component of normal pressure is negligible compared with the usual *tangential* stress involved in

viscosity. The present theory assumes stream line conditions. Any very large increase of $\frac{du_0}{dz}$ would probably introduce turbulence long before the normal pressure increase pa^2 would be measurable.

§ 26. *Conclusions.*

1. When Newtonian mechanics and elasticity theory are relied upon instead of probability, according to the geometrical weight method, the viscosity distribution function that results is

$$f=\frac{1}{a^3\pi^{3/2}}\exp\left(-\frac{c^2(1-p\cos^2\theta-q\sin^2\theta\cos^2(\phi-b))}{a^2\tan^2 b}\right). \quad (81)$$

$$f=\frac{1}{a^3\pi^{3/2}}\exp\left(-\frac{(u^2+v^2+w^2+a^2w^2(2x-1)^2-2auw(2x-1))}{a^2}\right),$$

where b, p, and q are related to the amount of shear $a(x)$ according to (20) and (85). The distribution (81) is characterized as follows :

(*a*) It is ellipsoidal in that the exponents, when set equal to the constant c'^2/a^2, are true strain ellipsoids as well as equal probability ellipsoids in velocity space, being obtained from the equal probability spheres of the Maxwell distribution by simple shear. The axes of symmetry of the ellipsoids are rotated ordinarily less than 1 per cent. of 1 second of arc from the position of spherical symmetry, and the eccentricities of the ellipsoids are each less than a thousandth.

(*b*) It comprises a single infinity of distributions, one for each group of molecules having traversed the x fractional part (between x and $x+dx$) of their several free paths since last previous impact.

(*c*) These distributions differ only by each one's characteristic amount of shear $a(2x-1)=\cot b -\tan b$, which varies linearly with x from $+a$ to $-a$ so as to make the distribution just before impact, when $x=1$, the mechanical image of the distribution after impact when $x=0$. And the distribution at the mid-points of the free paths where $x=\frac{1}{2}$ is the familiar Maxwell distribution. Numerically the amount of shear is ordinarily

less than one millionth and physically its constant "a" is the mean impressed shear per impact, the same for all velocities, according to a basic viscosity law of shear.

(*d*) It represents distortion, but no expansion in velocity space and the mean value of the amount of shear, for changes in x, is zero. And the entropy (strain) of the gas is accordingly the same as for equilibrium conditions at that temperature, because the impressed forces counterbalance the internal viscous stresses.

2. Boltzmann's continuity equation can be transformed so as to represent rate of change of range, strain, or entropy, and the distribution (81) is an exact solution, in conformity with two theorems applicable to the normal law for distributions and to equilibrium conditions in phase space.

3. The components of stress in velocity space (39) and (40) are related to each other in the same way as the components of stress (66), (73), and (74) in ordinary space, the component in the direction of the viscous force being greater by the factor $1+a^2$.

4. The simple pressure formula $p=1/3\ nm\bar{c}^2$, well known as the isothermal bulk modulus for gases obeying the gas law, and recently shown to be the elastic modulus for the equilibrium equations between stress and strain that yield the Maxwell distribution * and the Fermi-Dirac † distribution, appears again here as the elastic modulus (*a*) for the equilibrium equations (39) in velocity space that yield the ellipsoidal viscosity distribution function (81), and (*b*) for the equation $p_{rz}=pa$, which connects the shearing stress of viscosity with the amount of shear "a."

5. The viscosity coefficient is $\eta=\frac{\pi}{8}n\lambda m\bar{c}$.

6. The viscosity couple ‡ is $pa\lambda$, being proportional to the amount of shear and the absolute temperature.

The writer is much indebted to Prof. G. E. Uhlenbeck, of Ann Arbor, for very valuable criticism and discussions.

* Kimball, *loc. cit.*
† Kimball and Berry, *loc. cit.*
‡ Kimball and King, *loc. cit.*

II. *A Measurement of the Fundamental Sound generated by the Airscrew of an Aeroplane in Flight.* *By* E. T. Paris, *D.Sc., F.Inst.P.**

[Plate I.]

§ 1. *Introduction.*

THE object of the observations recorded in this paper was to determine the magnitude of the acoustical effects generated by the rotation of an airscrew on an aeroplane in flight.

The sound was recorded and its amplitude measured in C.G.S. units by means of a resonant hot-wire microphone which had been calibrated against a Rayleigh disk. Thus the method is one in which the tone to be examined is isolated by acoustical tuning, and the results may be of interest for comparison with those obtained by other methods, now generally in favour, in which the wave-form of the sound is recorded by an oscillograph and the energy of the constituent tones found by subsequent analysis.

The aeroplane available for the investigation was a Hawker Horsley torpedo carrier, with Rolls Royce Condor engine, driving a two-bladed wooden airscrew (diameter 4500 mm., pitch 3200 mm.) at about fourteen revolutions per second. It is well-known that a rotating airscrew generates a fundamental tone of frequency equal to the number of airscrew blades multiplied by the revolutions per second, and it was surmized at the time when the investigation was undertaken (1929) that this tone was physically the most important sound produced by an airscrew. This view has since been confirmed by the experiments of C. F. B. Kemp†, to which further reference is made in § 6.

§ 2. *The Sound-recording Apparatus.*

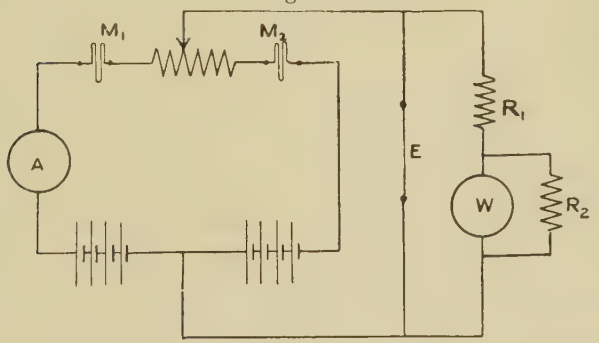
The sound-recording instrument was a tuned hot-wire microphone similar to that already described in connexion with observations on the distribution of sound round an airscrew ‡. The microphone was calibrated against

* Communicated by the Author.
† Proc. Phys. Soc. xliv. p. 151 (1932).
‡ Phil. Mag. xiii. p. 99 (1932).

a Rayleigh disk by the stationary-wave method, so that the actual pressure amplitude in the sound-wave could be deduced from the observed resistance changes. The theory of the calibration and some details of the apparatus are contained in § 4.

The microphone was used in a "battery bridge" circuit with a compensating grid as balancer (fig. 1). Two galvanometers were connected in parallel as shown in the figure. One was a Moll galvanometer with a period of about $\frac{1}{4}$ second, and the other an Einthoven string galvanometer. The values of the resistances R_1 and R_2 were adjustable; usual values were $R_1=400$ ohm, $R_2=0{\cdot}5$ ohm.

Fig. 1.



Electrical connexions to microphone.

M_1, active microphone; M_2 balancing microphone; A, milliammeter; E, string of Einthoven galvanometer; W, Moll galvanometer; R_1, R_1, resistances.

The Moll galvanometer was used to record the drop in the average resistance of the microphone grid, while the Einthoven galvanometer recorded the alternating resistance changes and was used for the accurate determination of the frequency of the sound.

The deflexion of the Moll glavanometer was recorded on a slow-running photographic recorder * carrying a strip of sensitized photographic paper 3·5 cm. wide, passing behind the slit of the camera at a rate of about $\frac{1}{8}$ inch per second.

* The recorder was the "sound-ranging" pattern made by the Cambridge Instrument Co.

The Einthoven galvanometer record was obtained in the following way :—The shadow of the string was thrown on to the slit of a high-speed recording camera of the type employed with the Low-Hilger audiometer. The shutter of the camera was adapted so that it could be operated electromagnetically, and an exposure could be made by the closing of a switch. When observations of the airscrew sound were being made the switch was worked by an observer in a *camera obscura* as described in § 3. The speed with which the sensitized paper moved behind the slit of the camera was about 60 cm./sec. By means of a " sound-ranging " time-wheel and tuning-fork time-markings were made on the photographic paper at intervals of 0·01 second. The length of paper on the drum of the Hilger camera was about 90 cm., so that an exposure lasting for about 1·5 sec. could be made.

The Einthoven record was required because the frequency of the sound could not be relied on to remain at a steady value throughout the time when observations were being made, and since variations in frequency produce changes in the steady drop in the resistance of the microphone grid (because the microphone is a tuned instrument) it was necessary to obtain a record of frequency simultaneously with the record of the Moll galvanometer deflexion. With an aeroplane in flight, apart from the possibility of variation in engine speed, the frequency of the sound arriving at the ground changes owing to the Doppler effect.

§ 3. *Method of Recording the Airscrew Sound from an Aeroplane in Flight.*

For recording the airscrew sound the microphone was placed at the middle of a rectangular concrete slab measuring 45 ft. × 30 ft. near the edge of an aerodrome. The surface of the concrete was roughly level with the surrounding ground, and the reflexion of the sound-waves from the concrete ensured that the pressure amplitude affecting the mircophone was exactly double the amplitude that would have existed at the same place if no reflexion had occurred. The photographic recording apparatus was in a laboratory at a considerable distance from the microphone.

Arrangements were made for the aeroplane to fly at a constant height of 1000 ft. on straight courses passing vertically, or nearly vertically, above the microphone, and observations of the amplitude of the fundamental airscrew sound emitted in the plane of rotation of the airscrew were made in the following manner:—The movements of the aeroplane were watched by an observer in a *camera obscura* about 25 yards from the concrete slab. This *camera obscura* had a lens of 30 in. focal length, and an image of the aeroplane could be seen on a horizontal map-board in the focal plane of the lens. The map-board could be rotated about a pivot which was vertically below the centre of the lens, and the board carried a sheet of paper on which were ruled a number of parallel "course lines," 1 inch apart, and also another line (the "crossing line") which was at right angles to the course-lines and passed through the central pivot. When an observation was to be made the observer in the camera watched for the appearance of the aeroplane's image in the map-board, and when it was seen the map-board was rotated until the image moved parallel to the course-lines. The observer was provided with a switch the closing of which lighted a small lamp near the slit of the sound-ranging recorder, and so caused a darkening of the film.

When a record was to be taken the observer waited until about half a second after the aeroplane had passed the crossing-line—to allow for the time taken by the sound to travel from the aeroplane to the microphone—and then closed the switch for about $1\frac{1}{2}$ seconds. The air-speed of the aeroplane being about 90 m.p.h. it was estimated that by this procedure the record of the airscrew sound was extended during the time the switch was closed from about 5° in front to 5° behind the plane of rotation. It was not, of course possible for the pilot to fly his machine exactly over the microphone on every occasion, but crossings in which the aeroplane did not pass within 6° of the vertical through the microphone were rejected. The distance from aeroplane to microphone could therefore be taken to be 1000 ft. in all cases. The length of track flown by the aeroplane on each crossing was about one mile— half a mile on each side of the microphone position.

Examples of the Moll galvanometer records are shown in figs. 2 and 3 (Pl. I.). The zero position is near the bottom

edge of the strip and a resistance change in the microphone grid of about 25 ohms was needed to deflect the trace across the strip. The closing of the switch by the observer also operated the shutter of the Low-Hilger camera and caused a photograph of the Einthoven string to be taken. Thus the dark band on each Moll galvanometer record indicates the time during which a frequency record was also being made. The duration of each band is approximately 1·5 seconds to correspond with the taking of the Einthoven record.

A reproduction of a portion of one of the Einthoven records is shown in Pl. I. fig. 4. The vertical markings divide the strip into hundredths of a second. In this particular record the frequency of vibration of the string was 58·8 cycle/sec., and since for sounds of large amplitude such as were observed in these experiments the frequency of the oscillatory resistance change in the hot-wire grid is twice the frequency of the aerial vibration affecting the microphone *, the frequency of the sound-waves was $\frac{1}{2} \times 58{\cdot}8 = 29{\cdot}4$ cycle/sec.

§ 4. *Calibration of the Hot-Wire Microphone.*

The hot-wire microphone was similar to that used in other experiments †, and the determination of the damping factor and conductance of the orifice have already been described ‡. The calibration was carried out by the stationary wave method, the stationary wave being generated in a pipe 21 ft. long and 2 ft. 6 in. in diameter. This length of pipe was needed in order to cope with the long wave-length of the airscrew fundamental; also a wide pipe had to be used, so that the microphone itself caused relatively little obstruction. In the stationary wave calibration a change δR in the resistance of the grid of the microphone was correlated with a relative pressure amplitude $\sin ky$, where $k = 2\pi/(\text{wave-length})$ and y is the distance the microphone was displaced from a loop in order to produce the resistance change δR. A Rayleigh disk suspended in the pipe served to indicate the amplitude of vibration at a loop, so that the pressure amplitude at a node could be calculated in dyne/sq. cm.

* *Cf.* Phil. Trans. A, cci. p. 410 (1921).
† Phil. Mag. xiii. p. 102 (1932).
‡ Proc. Phys. Soc. xliii. p. 74 (1931).

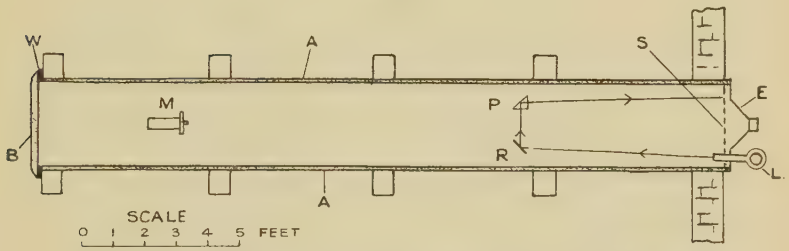
Thus if p_N dyne/cm.^{2} is the nodal pressure amplitude, then the amplitude producing the resistance change δR is $p_N \sin ky$ dyne/cm.^{2}, and so the microphone could be calibrated in C.G.S. units.

The arrangement of calibrating pipe, microphone, and Rayleigh disk is shown in fig. 5. The frequency of the sound used during a calibration was determined by a stroboscopic method *.

The Rayleigh disk was suspended at 45° to the direction of motion of particles in the sound-waves, so that the relation between the angular deflexion ϕ and the R.M.S. velocity v was, by König's formula †,

$$\phi = \frac{1}{6} \frac{\rho d^3 v^2}{S}, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (1)$$

Fig. 5.



Arrangement of microphone and Rayleigh disk in calibrating pipe.

A, concrete pipe ; B, concrete reflector ; W, cork washer ; M, Microphone ; R, Rayleigh disk (not to scale) ; P, prism (not to scale) ; S, translucent screen ; E, loud speaker ; L, lantern.

where ρ=density of air, d=diameter of disk, S=moment of torsion of the suspending fibre.

In the actual experiment the Rayleigh disk was not exactly at a loop, and a correction had to be applied on that account. Let Y represent the distance of the disk from the loop, then if V_L is the velocity amplitude at the loop the velocity amplitude at the disk is

$$V = V_L \cos kY. \quad . \quad . \quad . \quad . \quad . \quad (2)$$

Also, if a is the velocity of sound in air,

$$p_N = a\rho \, . \, V_L = a\rho \, . \, V/\cos ky. \quad . \quad . \quad . \quad (3)$$

* Proc. Phys. Soc. xliii. p. 76 (1931).
† Barnes & West, J. I. E. E. lxv. p. 873 (1927).

Since $v=V/\sqrt{2}$, (1) becomes

$$\phi=\frac{1}{12}\frac{\rho d^3 V^2}{S}, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (4)$$

so that

$$V=2\sqrt{\left(\frac{3\phi S}{\rho d^3}\right)}, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (5)$$

and

$$p_N=\frac{2a\rho}{\cos ky}\sqrt{\left(\frac{3\phi S}{\rho d^3}\right)}. \quad . \quad . \quad . \quad . \quad . \quad (6)$$

Or, if x be the observed deflexion of the spot of light on the translucent screen s (fig. 5) and l be the total optical length of path from the Rayleigh disk to the screen,

$$p_N=\frac{2a}{\cos ky}\sqrt{\left(\frac{3\rho x S}{ld^3}\right)}. \quad . \quad . \quad . \quad . \quad (7)$$

The disk used in the experiment was of silvered glass 0·519 cm. in diameter suspended on a glass fibre 29 cm. long. The value of S was found from the period, logarithmic decrement, and moment of inertia of a calibrating disk 1 cm. in diameter. It was 0·001870 erg/radian.

In the calibration of the microphone the Rayleigh disk was 321·6 cm. from the closed end of the pipe. The loop position was 276·1 cm. (one quarter-wavelength) from the closed end, and therefore $Y=321\cdot6-276\cdot1=45\cdot5$ cm. The total optical path from the Rayleigh disk to the screen was 320·3 cm. The supply to the loud-speaker was adjusted until the output was strong enough to give a deflexion (x) which could be read with accuracy, namely, 37·2 cm. The temperature in the pipe was 11°·3 C., so that $a=33830$ cm./sec. The barometric pressure was 751 mm., so that $\rho=0\cdot001224$ gm./c.c. The frequency of the sound was 30·63 cycle/sec., whence $k=0\cdot326$ degree/cm. Hence, using the value of S previously given, we find from (8) that the pressure amplitude at the node was

$$p_K=167 \text{ dyne/cm.}^2.$$

It was found that the strength of the sound in the calibrating pipe when adjusted to suit the sensitiveness of the Rayleigh disk was too great for the calibration of the resonant microphone, the difficulty being that the microphone had only to be moved a few centimetres

from the loop position for the change of resistance in the grid to rise to 10 or 15 ohms, so that the calibration could not be effected with any precision. To overcome this difficulty a smaller container was used for the calibration, the "neck" of the original microphone (which contains the grid) being removed and placed on the smaller container. The frequency of the microphone during calibration was thus raised to about 38 cycle/sec. (the exact value depending on air temperature) as compared with 29 cycle/sec. when used to record the airscrew.

An allowance for this change of frequency during calibration was made as follows :—Suppose that a change of resistance δR is observed when the aeroplane sound affects the recording microphone. This corresponds to some definite amplitude of oscillation in the neck of the microphone—that is, to some definite value of the volume displacement of q. Now q is related to the velocity potential ψ by

$$q=\Omega\psi, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (8)$$

and therefore to the pressure variation p by

$$q=\Omega p/\iota\omega\rho, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (9)$$

where

$$\Omega=\frac{c}{2h-\iota(\omega_0{}^2-\omega^2)/\omega} \quad . \quad . \quad . \quad . \quad (10)$$

c being the conductance of the neck of the microphone resonator and h its damping factor. In the calibration the same resistance change δR is produced by a known pressure variation P at a frequency $\omega'/2\pi$ (which differs slightly from $\omega/2\pi$, the frequency of the recorded airscrew sound), while the natural frequency of the microphone is changed from $\omega_1/2\pi$ to some other value $\omega_1'/2\pi$. Thus we may write

$$q=\Omega' P/\iota\omega'\rho, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (11)$$

where

$$\Omega'=\frac{c}{2h-\iota(\omega_0{}'^2-\omega'^2)/\omega'}. \quad . \quad . \quad . \quad . \quad (12)$$

The values of c and h are of course the same because the same orifice is used. The value of h is influenced by frequency, but since ω and ω' were very slightly different

(within 1 cycle/sec. on an average) the difference may be ignored. Thus from (9) and (11)

$$\frac{\Omega p}{\omega} = \frac{\Omega' \mathrm{P}}{\omega'}, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (13)$$

and

$$p = \frac{\omega \Omega'}{\omega' \Omega} \mathrm{P}. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (14)$$

From the calibration, Ω'/ω' and P were known, and it was only necessary therefore to determine Ω/ω for each record of airscrew sound obtained.

The procedure actually adopted was to measure the photographic records and deduce the mean ohmic change δR in the microphone. From the calibration in the pipe the distance y which it would have been necessary to move the microphone from the loop in order to produce this resistance change δR was determined. If p_N is the pressure variation at a node in the calibration pipe, we see that

$$\mathrm{P} = p_N \sin ky, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (15)$$

whence

$$p = \frac{\omega \Omega'}{\omega' \Omega} p_N \sin ky. \quad . \quad . \quad . \quad . \quad . \quad (16)$$

§ 5. *Record of Observations.*

Typical records, obtained with the Moll galvanometer, from which the resistance change in the microphone grid was determined are reproduced in Pl. I. figs. 2 & 3. The cause of the irregularities in the records is not known with certainty, but fluctuations in the acoustic output of the airscrew and irregularities in the structure of the atmosphere are probably sufficient to account for them.

The following table shows the results obtained on the two days on which observations were made. The value of the resistance change δR was deduced from the mean value of the deflexion during the darkened part of the record, when a frequency record was being made simultaneously.

The values of p in the last column of the table are the pressure amplitudes at the reflecting concrete surface on which the microphone was laid. The mean value is 3·9 dyne/cm.^{2}.

In the absence of reflexion the pressure amplitude would be halved and its mean value would be 1·95 dyne/cm.^{2}.

Table of Observations.

Date.	Record no.	Frequency of sound.	δR.	sin *ky*.	$p_N\left(\frac{\omega\Omega'}{\omega'\Omega}\right)$.	*p*.
		cycle/sec.	ohm.		dyne/cm.^{2}.	dyne/cm.^{2}.
20.2.29	2	29·4	10·9	0·1138	30·8	3·5
	3	29·25	13·7	0·1305	30·3	4·0
	4	30·4	14·4	0·1334	36·2	4·8
	5	30·3	12·4	0·1234	35·7	4·4
21.2.29	2	29·1	10·65	0·1129	29·7	3·4
	3	28·6	11·6	0·1187	29·8	3·5
	4	29·0	16·3	0·1429	29·7	4·2
	5	28·4	13·2	0·1282	30·2	3·9
	6	29·3	13·3	0·1288	30·3	3·9
	7	28·8	12·8	0·1259	29·7	3·7
	8	29·5	11·95	0·1204	30·6	3·7

§ 6. *The Acoustic Output at the Fundamental Frequency.*

At 1000 ft. from the airscrew the energy flux in the sound-waves may be calculated in the same way as for plane-waves. Thus the flux associated with a pressure amplitude is

$$E = p^2/2R \text{ erg/cm.}^2\text{/sec.}, \quad . \quad . \quad . \quad . \quad (17)$$

where R is the acoustic resistance of air equal to 41·9 gm./cm.^{2}/sec.

If polar coordinates r, β, ϕ be used, where β is measured from the axis of rotation (in the direction of flight) and ϕ from some selected direction in the plane of rotation, the total output will be

$$W = \int_0^{\pi} \int_0^{2\pi} E_\beta r^2 \sin\beta \, d\beta \, d\phi, \quad . \quad . \quad . \quad (18)$$

where E_β is the energy flux in the direction β at distance r. Since there is symmetry round the axis of rotation

$$W = 2\pi r^2 \int_0^{\pi} E_\beta \sin\beta \, d\beta, \quad . \quad . \quad . \quad . \quad (19)$$

or from (17)
$$W = (\pi/R) \int_0^{\pi} (rp)^2 \sin\beta \, d\beta$$
$$= 0{\cdot}075 \int_0^{\pi} (rp)^2 \sin\beta \, d\beta \quad , \quad . \quad . \quad . \quad (20)$$

p being the pressure amplitude in the direction β at distance r.

It will be assumed that the variation of p with β is the same as when the airscrew is rotating at zero rate of advance. Experiments to determine this variation have been made by C. F. B. Kemp * and by the writer †, but the results are not in complete agreement. Although the polar curves exhibiting the variation of p with β agree in their salient features (namely, that p is minimal when $\beta=0$ or π, that there is a principal maximum of p when β is about 120°, and a secondary maximum when p is about 30°), Kemp's curve indicates a relatively greater output in the direction $\beta=90°$. Consequently the calculated output for the airscrew in flight depends on which curve is assumed to be correct. The two calculated results are 29 watts (Kemp) and 55 watts (Paris).

Both these results are considerably greater than the output found by Kemp for an airscrew at zero rate of advance, namely 18 watts ‡.

A factor which affects the comparison between the output in flight and at zero rate of advance is the rotational speed of the airscrew. In flight the average engine speed was 1850 r.p.m., corresponding to a fundamental airscrew frequency of 29 cycle/sec. (the ratio of airscrew speed to engine speed being 0·477), whereas at zero rate of advance the engine speed was 1750 r.p.m. It is known that there is considerable increase of amplitude with rotational speed, and part at least of the excess in the calculated output in flight over that at zero rate of advance could be accounted for in this way.

§ 7. *Summary.*

A measurement of the amplitude of the fundamental tone generated by the airscrew of an aeroplane in flight is described. A resonant hot-wire microphone, calibrated against a Rayleigh disk, was employed, the sound being observed on the ground directly below the aeroplane

* *Loc. cit. suprà.*

† Phil. Mag. xiii. pp. 105–106 (1932).

‡ The outputs given in Kemp's table 2 (p. 158, *loc. cit.*) are not consistent with the values of p indicated in his figure 2 (p. 157). I am informed by Mr. Kemp that through an error in calculation the outputs in table 2 should be doubled. Thus for the fundamental frequency the output should be $2\times 8{\cdot}9=18$ watts approximately.

flying at 1000 ft. An estimate is made of the acoustic output of the airscrew at the fundamental frequency. The output in flight appears to be greater than that at zero rate of advance.

III. *The Propagation of Sound along the Slipstream of an Airscrew. By* E. T. Paris, *D.Sc., F.Inst.P.**

DURING an investigation relating to the nature and distribution of aeroplane sound some observations were made on the distribution of the second harmonic of the exhaust sound of a Rolls Royce Condor engine. These observations, though only of a qualitative nature, indicated that the sound propagated along the slipstream of the airscrew, in the direction of motion of the air, became rapidly attenuated.

The engine was in a Berkeley fuselage from which the wings, tail-plane, and rudder had been removed. The fuselage was on an open space of ground and lashed down to screw pickets, with its axis horizontal, so that the airscrew could rotate in a vertical plane †.

The Condor is a twelve-cylinder V-type engine, and each bank of six cylinders exhausted through a separate pipe about 4 in. in diameter, with an outlet just below the fuselage and about 3 ft. above the ground. The two outlets were 46 in. apart and placed symmetrically on each side of the median plane of the fuselage. The outlets lay within the slipstream of the airscrew (14 ft. 9 in. in diameter) and about 5 ft. below the axis of rotation.

The theory of the exhaust sound from a twelve-cylinder V-type engine has been given by M. D. Hart‡. Each exhaust outlet acts as an independent source of sound of fundamental frequency $\frac{1}{2}$NR, where N is the number of cylinders in each bank and R is the revolutions of the engine per second. The two sources are out of phase by half the period of the fundamental, and thus form a doublet of fundamental frequency $\frac{1}{2}$NR. The intensity

* Communicated by the Author.

† The properties of the airscrew sound have been described by C. F. B. Kemp, Proc. Phys. Soc. lxiv. p. 153 (1932).

‡ " The Aeroplane as a Source of Sound," Aeron. Res. Comm. R & M., no. 1310, pp. 7–10 (1929).

or energy of the qth harmonic, in a direction making an angle β with a horizontal line at right angles to the axis of the doublet, is proportional to

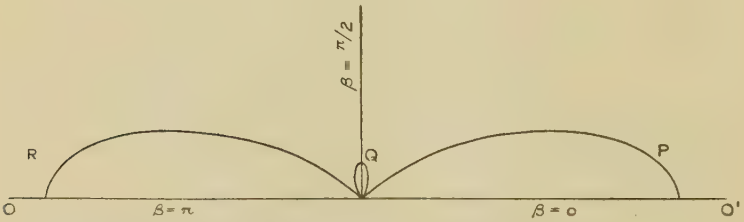
$$\left\{1+(-1)^q \cos \frac{\pi q l \mathrm{RN} \sin \beta}{a}\right\}, \quad . \quad . \quad . \quad (1)$$

where l is the distance apart of the sources or outlets and a is the velocity of sound. For the second harmonic ($q=2$), with $l=46$ in., R$=1750/60$ r.p.s., and N$=6$, the theoretical distribution of intensity is given by

$$\{1+\cos (1{\cdot}22\pi \sin \beta)\}. \quad . \quad . \quad . \quad . \quad . \quad (2)$$

The polar curve corresponding to this expression is reproduced in fig. 1. The principal points to be noted

Fig. 1.



Theoretical distribution of second harmonic of exhaust sound from twelve-cylinder V-engine.

are that most of the sound is projected forward in the direction $\beta=0$ or backwards in the direction $\beta=\pi$, and that there is symmetry fore and aft of the engine—that is, about the line $\beta=\pi/2$. The curve also has a small lateral lobe with a maximum along $\beta=\pi/2$. The complete curve is of course symmetrical about the horizontal axis 00′.

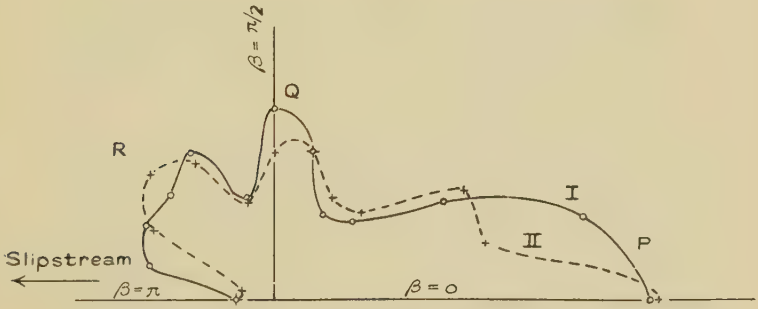
The distribution of sound round the engine was explored with the aid of a hot-wire microphone tuned to the second harmonic of the exhaust. The microphone was placed successively in thirteen positions on the circumference of a semicircle of 100 yards radius with the engine at the centre, the angular separation of adjacent positions being 15°. Two sets of observations (curves I. and II.) are shown graphically in fig. 2, in which the deflexions of a galvanometer (proportional to the resistance-change in the microphone grid) are plotted against β in a polar

diagram. The fuselage with the engine was turned "end for end" between the two sets of observations so as to ensure that the observed distribution was not unduly influenced by reflexion from surrounding objects.

The observations show that instead of large outputs both forward and backwards, as indicated by theory, the sound-energy is greatest in the forward direction ($\beta=0$), while the amount of sound sent backwards along the slipstream, *i. e.*, in the direction $\beta=\pi$, is relatively small.

The microphone used in these observations was not calibrated, and therefore a quantitative comparison between the theoretical and experimental curves cannot

Fig. 2.



Observed distribution of second harmonic of exhaust sound from twelve-cylinder V-engine.

be made. The resistance-change (δR) in a hot-wire microphone of the type employed, in the range 0 to 30 ohms (where the present observations lay), is connected with the pressure-amplitude (p) in the sound-wave by a relation of the form $\delta R = \text{constant} \times p^n$, where n is greater than 2. Usual values for n are from 2·1 to 2·6. Hence it can at least be said that the observed output of sound laterally in the neighbourhood of $\beta=\pi/2$ is greater than that indicated by theory.

The maxima at P, Q, and R in the theoretical curve in fig. 1 may be identified with those at P, Q, and R in the experimental curves in fig. 2. It appears that the effect of the slipstream is to divert sound-energy from directions near to $\beta=\pi$, and to send it instead in the lateral directions Q and R in fig. 2.

The air-velocities in the slipstream in which the exhaust outlets lay may be of interest. These were measured with an air-meter, and it was found that at the same level as the outlets and 26 ft. behind them, along the axis of the fuselage, the velocity was 95 ft./sec. At a distance of 6 ft. from the axis it was 70 ft./sec., while at 12 ft. from the axis it had dropped to 10 ft./sec. Along the axis and 72 ft. from the exhaust outlets the velocity was 78 ft./sec. ; at 117 ft. it had fallen to 60 ft./sec., and at 327 ft. to 25 ft./sec.

IV. *The Electrical Conductivities of Aqueous Solutions of Sodium and Potassium Hydroxides at 25° and the Validity of the "Water Correction."* *By* George Harold Jeffery, *B.Sc.(Lond.), A.I.C., and* Arthur Israel Vogel, *D.Sc.(Lond.), D.I.C., F.I.C.**

THE criticism has been made by a number of eminent chemists that our measurements upon the electrical conductivities of sodium and potassium hydroxide solutions † at 25° are subject to considerable error owing to our application of the water correction on the assumption that carbonic acid is the only impurity, it being suggested that small quantities of dissolved salts and other impurities were present in the equilibrium water used. Independent measurements by Goworecka and Hlasko ‡, which were published whilst our paper was in the press, afford confirmation of our final "corrected" results and therefore provide experimental evidence in support of our method of application of the solvent correction. These workers carried out two series of determinations, the first with ultra-pure conductivity water of specific conductivity of about 0·1 gemmho in which the solvent correction is almost negligible, and the second with ordinary conductivity water to which a slight excess of dilute barium hydroxide solution was added to precipitate all the carbonic acid and due allowance was made for the presence of the

* Communicated by the Authors.
† Phil. Mag. (7) xv. p. 395 (1933).
‡ *Recz. Chem.* xii. p. 403 (1932).

excess of barium hydroxide upon the conductivities of the alkali hydroxides. The results obtained from both series were in excellent agreement. A comparison between Goworecka and Hlasko's figures [G. and H.] and our own "corrected" results [J. and V.] at round concentrations, which were read off from a conductivity-concentration curve drawn with a flexible wooden spline, is shown in the following table :—

	Sodium hydroxide.		Potassium hydroxide.	
$C \times 10^4$.	G. and H.	J. and V.	G. and H.	J. and V.
5....	246·1	247·2	269·5	270·8
10....	244·7	245·1	269·1	268·7
20....	242·8	242·7	266·2	266·4
30....	241·5	241·3	265·0	265·0
40....	240·7	240·0	263·9	263·8
50....	239·9	238·8	262·5	262·8
60....	239·2	237·9	262·0	261·9
70....	238·5	236·9	261·5	261·0
80....	237·9	236·1	261·0	260·2
90....	237·4	235·3	260·5	259·5
100....	237·0	234·5	260·0	258·9

The agreement up to ca. 0·005 N., over which range the effect of the water correction is the most marked, is within 0·4 per cent. for sodium hydroxide and 0·1 per cent. for potassium hydroxide. Larger discrepancies are apparent in the more concentrated solutions, but these may be due to the fact that these concentrations are outside the effective concentration ranges of the conductivity cells. Unfortunately, Goworecka and Hlasko do not give sufficient details of the construction of their cells to test this view. It appears, therefore, that in dilute solutions, prepared with equilibrium water *, the only impurity which need be taken into account is carbonic acid.

* Vogel and Jeffery, J. Chem. Soc. (1931) p. 1201.

Woolwich Polytechnic, London, S.E. 18,
and University College, Southampton.

V. *The Electrocapillary Maximum for a Mercury Electrode in contact with a Solution containing Mercurous Ions.* *By* S. R. Craxford *.

Introduction.

ELECTROCAPILLARY curves are normally obtained by polarizing a mercury-aqueous solution interphase, and have their maximum surface tension in the region of $\epsilon_H = -0{\cdot}25$ volt on the normal hydrogen scale, a value that is independent of whether the surface tension is measured by the capillary-rise method in a capillary electrometer, or by a drop-weight method. As early as 1897, however, Nernst [1] suggested, and demonstrated qualitatively, that similar curves could be obtained if the potential across the interphase were fixed by the mercurous ion concentration in the solution instead of by polarization, but it is only recently that quantitative examples have been published. Bennewitz and Delijannis [2] obtained such a curve by measuring the surface tension by the drop-weight method, and it was remarkable in having its maximum at $\epsilon_H = +0{\cdot}48$ volt. Their method consisted essentially in dropping mercury from a capillary tip into a solution containing mercurous salt, and the potential of the dropping mercury was assumed to be the same as that of a flat stationary mercury electrode in the same solution. The potential of the latter was altered as required by varying the concentration of the mercurous salt, and was measured directly against a standard electrode. But the concentration of mercurous salt, from 10^{-1} down to 10^{-13} normal, was excessively small for the greater part of the curve, so that the rate of attainment of thermodynamic, *i. e.*, statistical, equilibrium between mercurous ions on either side of the interphase was low, thus making it reasonable to suppose that in such cases the potential of the expanding drop is no longer completely determined by the mercurous ion concentration in the solution. As this concentration decreases, the surface of the mercury drop tends to become more and more polarizable, and its potential reaches the true electrocapillary maximum at $\epsilon_H = -0{\cdot}25$ volt if the discharge of

* Communicated by O. Gatty, B.Sc.

other ions from solution is negligible. In the general case, the potential of the drop will be determined by the relative rates of expansion and of attainment of electrochemical equilibrium. This point of view is confirmed by Rosenberg and Stegeman [3], who measured the potential of dropping mercury in solutions of mercurous salts of various concentrations.

On similar grounds Frumkin and Obrutschewa [4] pointed out that the error in Bennewitz and Delijannis's work lay in the assumption that the stationary and the dropping mercury were at the same potential, and therefore, in order to obtain correct electrocapillary curves the dropping and stationary mercury electrodes must be connected electrically to bring them both to the same potential. In this way they obtained an electrocapillary curve for N/10 KNO_3 and found that it was superposable on the corresponding part of the electrocapillary curve obtained by polarizing the interphase, and that it showed no trace of a maximum near $\epsilon_H = +0{\cdot}48$ volt.

Bennewitz and Küchler, however, extended the previous work and found that in all cases the curves are altered when the dropping mercury electrodes are connected. But while agreeing that for solutions containing much potassium nitrate the corrected curve approximates to the polarized electrocapillary curve, they show that for solutions containing mercurous nitrate and a trace of nitric acid only, the corrected curve still has a maximum in the positive region, but this time at $\epsilon_H = +0{\cdot}52$ volt. The solutions used to give these results contained only a very small trace of mercurous salt, and so must have had a very large electrical resistance. Moreover, the present work shows that in the region of their maximum a considerable current must flow from the stationary to the dropping electrode to maintain the latter at the correct potential, so that there may be a large ohmic drop of potential across the high resistance of the cell, and so the potential across the dropping interphase may again become uncertain. Thus the only way to obtain results that may be interpreted without fear of error, is to measure the potential of the dropping mercury directly. The results of such measurements are given in the present paper, and in no case do they show any trace of a maximum near $\epsilon_H = +0{\cdot}5$ volt. Further discussion of these results is postponed until a later section.

The Drop-Weight Method for determining Electro-Capillary Curves.

The exact theory of the drop-weight method of measuring surface tension is still uncertain, but it is generally agreed that results are expressed to a good degree of approximation by the equation

$$\gamma = \frac{mg}{r} \cdot \mathrm{F},$$

where

γ is the surface tension,

mg is the weight of the drop,

r is the radius of the orifice where it is formed,

F is a function of V/r^3,

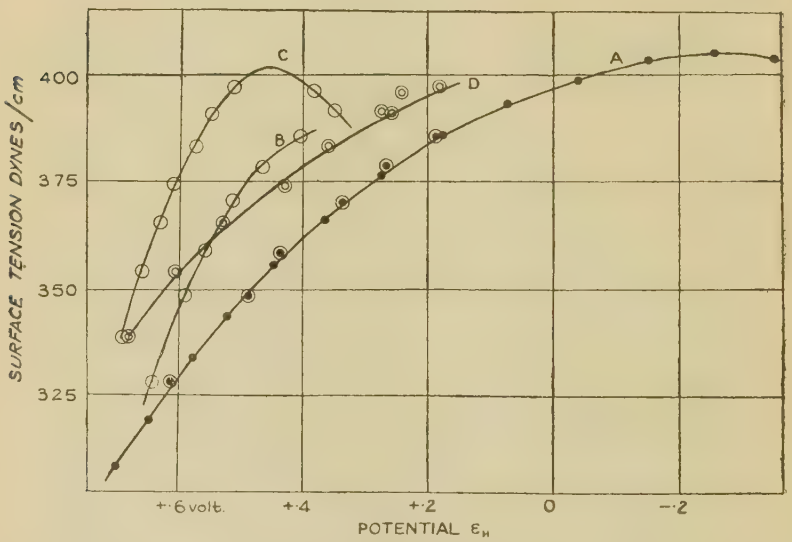
V is the volume of the drop.

Thus the function F must be known, not only if the surface tension is to be calculated absolutely, but also if two surface tensions are to be compared by the drop-weight method. Tables of values of the function F for various values of V/r^3 are given by Harkins and Brown [6], who compared the drop weights with the surface tensions measured by the capillary-rise method for the interphases water-benzene, water-carbon tetrachloride, and water-ethylene dibromide, and it is customary to apply the corrections given by these tables in all cases. Thus, in considering the relative merits of the drop-weight and the capillary-electrometer methods for the determination of electrocapillary curves, it must be counted to the disadvantage of the drop-weight method that it involves the use of a correction term derived from measurements of a water-benzene interphase, and which has not been proved directly to be applicable to a water-mercury interphase. This uncertainty in calculating the ratios of surface tensions from ratios of drop weights counterbalances any superiority the drop-weight method may have over the capillary-electrometer method for determining electrocapillary curves, by virtue of the fact that the interphase of an expanding drop will be relatively free from contamination by products of electrolysis and will be surrounded by a layer of solution which is constantly being renewed, and so does not tend to become exhausted of

one or other of its components by electrolysis. If electrocapillary curves drawn by the two methods were compared it would be impossible to decide how much of the discrepancy, if any, would be due to one cause, and how much to the other. Kučera [7] compared electrocapillary curves measured by both methods and found rough agreement if the surface tension is taken simply as proportional to the drop weight. This is done in the present paper, and is probably correct if the elementary precautions described by Kučera are observed, namely, if the apparatus is protected from mechanical disturbances and the rate of dropping is kept approximately constant throughout the work.

The apparatus used is practically as described by Kučera. A mercury reservoir is connected by a rubber tube to a glass capillary tube from the lower end of which the drops fall. Before use this rubber tube is boiled out with caustic soda and then repeatedly with water, until the latter remains clear. It is then dried and allowed to stand in contact with mercury for some time to remove any remaining trace of impurity that might attack mercury. The capillary is a piece of thermometer tubing, of internal radius 0·012 cm., broken off cleanly at the lower end and constricted in the middle, so that, with a pressure head of 20 cm. of mercury, drops are only formed at the tip at a rate of one every 3 or 4 seconds. For each measurement the pressure head is adjusted to give a dropping rate of one drop in 3·65 seconds approximately. The solution to be investigated is contained in a beaker, which also contains a layer of mercury to act as an electrode, and which is connected through an intermediate vessel to a saturated calomel reference electrode. All the glass parts are initially cleaned in a mixture of nitric and chromic acids and washed out thoroughly with distilled water, and finally with conductivity water. The mercury used was previously treated with dilute nitric acid in the apparatus described by Desha [8], and was then distilled in a Hullet still so as to eliminate as far as possible other metallic impurities. The beaker is first filled with either a solution of N/10KNO_3 or with pure water as required, and the potential of the layer of mercury in the bottom adjusted by the method of Frumkin and Obrutschewa, namely, by adding small amounts of a solution of mercurous nitrate on the one hand, or traces of alkali on the

other. All solutions were prepared from pure recrystallized salts and conductivity water from which dissolved air had previously been removed. The mercurous nitrate solutions had present an equal molal concentration of nitric acid to keep the mercurous salt in solution. The mercury is allowed to start dropping from the capillary tip before the latter is inserted into the solution, so as to avoid wetting the interior of the capillary. For every point on the curve the appropriate potentials were measured, then 50 drops were collected in a glass spoon and removed, and the potentials again measured. Three lots



A ● The polarized curve for N/10 KNO_3.

B ◯ The non-polarized curve for N/10 $KNO_3 + Hg_2(NO_3)_2$, using the potential of the stationary mercury electrode.

◯ As above, but using the potential of the dropping mercury.

C ◯ Non-polarized curve for $Hg_2(NO_3)_2$ only, using the potential of the stationary mercury.

D ◎ As C, but using the potential of the dropping mercury.

of 50 drops were removed in this way for each point, and each lot of mercury was dried, first with filter paper and then in a vacuum desiccator over phosphorus pentoxide, and finally weighed. Working in this way the

standard deviation in the weight of 50 drops was 0·00045 gm. in those cases where the conditions in the solution could be controlled.

Experimental Results.

The experimental results are shown in the figure. Curve A is the usual electrocapillary curve obtained by polarization for a solution of N/10 potassium nitrate, and

TABLE I.

Electrocapillary curve for N/10 KNO_3, the interphase being polarized.

ϵ_H volts.	γ dynes/cm.
+0·699	309
+0·646	320
+0·578	334
+0·520	344
+0·446	355
+0·365	364
+0·277	376
+0·176	385
+0·071	393
−0·039	399
−0·152	403
−0·257	404
−0·353	403

is drawn from the data in Table I. The negative branch of the curve is omitted as it is not of interest for the present purpose. The drop weight at the maximum is compared with Gouy's [9] value of the surface tension of 427 dynes/cm. and this is used to convert all other drop-weight readings into surface tensions in dynes/cm., on which basis all the curves are plotted. Curve B is obtained for the same solution when the potential is adjusted by the addition of mercurous nitrate instead of by polarization, and the surface tension is plotted against the potential of the flat mercury electrode in the manner of Bennewitz and Delijannis. The points shown in double circles, which fall back on to the first electrocapillary curve, are obtained

when the surface tension as measured for curve B is plotted against the potential of the dropping mercury measured directly. These results are shown in Table II. The observed potential fluctuates as the individual drops form and break away, but by watching the drop while the potential is being measured on a potentiometer, it is possible after a little practice to obtain the potential reached just before the drop breaks away. This clearly is the potential which determines the surface tension. The fact that these points fall back on to the electrocapillary curve obtained by polarization confirms the work of Frumkin and Obrutschewa, in which similar points

TABLE II.

Electrocapillary curve for N/10 KNO_3, the potential being altered by the addition of $Hg_2(NO_3)_2$.

γ dynes/cm.	ϵ_H volts (stationary electrode).	ϵ_H volts (dropping electrode).
328	+0·644	+0·615
349	+0·590	+0·488
359	+0·559	+0·438
371	+0·515	+0·338
378	+0·465	+0·267
385	+0·406	+0·186

were brought back on to the normal electrocapillary curve by short circuiting the dropping and the stationary mercury electrodes. This means that, as Nernst had previously supposed, it is immaterial whether the potential across the mercury-aqueous solution interphase is fixed by polarization or by the composition of the solution.

Curve C is that obtained by plotting the surface tension against the potential of the flat mercury electrode for solutions containing mercurous salt only (Table III.). These solutions, being excessively dilute, were only very feebly conducting. This curve is of the type obtained by Bennewitz and his co-workers with a maximum near $\epsilon_H = +0{\cdot}48$ volt. In the region of this maximum the surface tension and potential readings are neither steady nor reproducible, as they are for the rest of the ascending branch of the curve, and this is not very surprising, since

these quantities are determined by the very small concentrations of substances present in the solution, which concentrations are not susceptible to direct experimental control. When, however, the variation of surface tension with the potential of the dropping mercury is considered, as in curve D, there is no trace of any such maximum and the curve appears to be a portion of a normal electrocapillary curve. Compared with the electrocapillary curve for N/10 potassium nitrate this one is displaced consider-

TABLE III.

Electrocapillary curve for water, the potential being altered by the addition of $Hg_2(NO_3)_2$.

γ dynes/cm.	ϵ_H volts (stationary mercury).	ϵ_H volts (dropping mercury).
339	+0·692	+0·680
354	+0·660	+0·607
366	+0·634	+0·532
375	+0·611	+0·431
383	+0·576	+0·362
391	+0·550	+0·260
397	+0·513	+0·182
396	+0·383	+0·244
392	+0·351	+0·276

ably towards the positive side of the diagram, a result which is readily explained by Stern's [10] theory of the electrolytic double layer, according to which, curves for very dilute solutions should be so displaced.

These results show clearly that that there is no true electrocapillary maximum in the region of $\epsilon_H = +0.5$ volt, as had previously been claimed. Further discussion of the implications of this work is undertaken in the next section.

The Absolute Zero of Potential.

In order to avoid the philosophical difficulty raised by Guggenheim [11] as to the meaning of the electrical potential difference between two dissimilar phases, it may be stated at once that all the methods that claim to measure

the absolute zero of electrolytic potential really measure the potential at which there is no net charge on either phase. This means that the net charge in the interphase on either side of a boundary called the phase boundary, is zero, and such a point will be defined as the electrolytic null-point. If there is a charge separation near the interphase, but confined to one phase only, due to a cause such as the presence of a layer of oriented dipoles, or to specific adsorption, there will still be a difference of potential between the two phases at the null-point. This is manifested by the fact that the normal electrocapillary maxima occur at more negative potentials when there are present in the solution ions that are specifically adsorbed at the interphase. But even in extreme cases the shift does not exceed 0·4 volt (see Gouy, *loc. cit.*).

If the results of methods depending on electrophoresis be passed over, as the theory of electrophoretic potentials in relation to interfacial potential differences is still obscure, the remaining methods give values for the potential of the null point which fall into two sharply conflicting classes. Methods of class A give values in the region of $\epsilon_H = -0{\cdot}25$ volt and include the method of the electrocapillary maximum and of the dropping mercury electrode. When the interfacial potential difference is fixed by polarization, the null-point is given this value both by the electrocapillary maximum (for the best curve see Koenig [12]), and by the dropping electrode as used by Philpot [13] and also as used by Paschen [14], when it was arranged so that the chances of electrochemical equilibrium occurring were minimised, and then the dropping mercury came to the potential of the null-point. On the other hand, methods of class B, which give the potential of the null-point near $\epsilon_H = +0{\cdot}5$ volt, are the scrapped electrode method of Bennewitz and Schulz [15] and of Bennewitz and Bigalke [16]; the electrocapillary maximum method of Bennewitz and his co-workers, which has been described in this paper, and the dropping electrode method of Billiter [17], in which the potential of a dropping mercury electrode was assumed to be controlled by the composition of the solution, and then it was determined whether the dropping mercury adsorbed mercurous salt from, or gave it up to, the solution.

The methods of class B depend on the assumption that

the potential of an expanding interphase is determined by the composition of the solution, that is, by complete thermodynamic equilibrium between the two phases, and hence, that the potential difference across such an expanding interphase is the same as that across a stationary one in the same solution. This, however, is untrue, and the work of Rosenberg and Stegeman for rapidly dropping mercury, confirmed by the present work for slower rates of dropping, proves conclusively that the dropping mercury in Billiter's experiments must have been at a potential far removed from the potential he assumed it to have. Thus his results must be in error, probably to the extent of 0·5 volt, and so need not be considered as giving any value for the null-point.

The experimental part of the present paper described in the last section proves the same to be true for the electrocapillary maxima said to occur in the region $\epsilon_H = +0{\cdot}5$ volt. Thus at the moment the balance of evidence is very strongly in favour of the null-point having a potential near $\epsilon_H = -0{\cdot}25$ volt.

The Theory of Electro-Capillarity.

In a previous section it was pointed out that it is immaterial how the potential across a mercury-solution interphase is fixed, whether by polarization or by the composition of the solution. Thus there are two cases, thermodynamically and physically very different, which lead to the same result, and it is proposed to consider the theory of both cases to see what deductions may be made from the identity of the final results.

In the case of the polarized interphase the fundamental phenomenon is that if the area of the interphase is increased by ds, a charge dq must flow round the external circuit if the potential ψ across the interphase is to be retained constant. In this process mechanical work $\gamma\, ds$ is required, and electrical work $\phi\, dq$, where ϕ is the potential difference between the terminals of the apparatus used, and dq is positive when it flows against ϕ. Then it may be shown by thermodynamical reasoning [23] that

$$\left(\frac{\partial \gamma}{\partial \phi}\right)_s = -\left(\frac{\partial q}{\partial s}\right)_\phi .$$

But as the electrode in the solution is taken to have constant potential, while the potential ψ of the interphase we are considering varies, $d\phi = d\psi$, so that

$$\left(\frac{\partial\gamma}{\partial\psi}\right)_s = -\left(\frac{\partial q}{\partial s}\right)_\psi \quad . \quad . \quad . \quad . \quad . \quad . \quad (1)$$

This is essentially the equation derived by Lippmann [18] in 1873. The right-hand side was interpreted in a purely electrostatic manner by Helmholtz [19], who considered $\left(\frac{\partial q}{\partial s}\right)_\psi$ to be the charge Q per unit area of the electrolytic double layer, as defined in the last section. The capacity per unit surface area, C, is $\frac{d\mathrm{Q}}{d\psi}$, and if the double layer is considered, with Helmholtz, to be analogous to a parallel plate condenser, then $\mathrm{C} = \mathrm{Q}/\psi$, and the Lippmann equation becomes

$$\frac{d\gamma}{d\psi} = +\mathrm{C}(\psi - \psi_0),$$

the equation of a parabola. Since Q has been proved not to be strictly a linear function of ψ this simple picture of the double layer is not quite correct, and as a result the whole theory fell into disrepute. But up to and including the statement that

$$\left(\frac{\partial\gamma}{\partial\psi}\right)_s = -\mathrm{Q}$$

it is correct for a completely polarizable interphase, and a direct experimental verification is given by Frumkin [24], who measured Q with a dropping mercury electrode at a series of potentials and compared it with the slope of the corresponding electrocapillary curve at the same potentials.

The theory of the non-polarized case is merely an extension of the Gibbs adsorption equation $\frac{d\gamma}{d\mu_x} = \Gamma_x$, where μ_x is the chemical potential of the species x in the solution, and Γ_x is the amount of this component adsorbed per unit area in the interphase. The precise theory of this case is being published in another paper [23], but the rather less exact treatment that has been used

up to the present is sufficient to illustrate the argument of this section. Thus by Nernst's theory of electrode potential

$$n\mathrm{F}d\psi = d\mu_{\dot{\mathrm{H}}\dot{\mathrm{g}}_2}$$

we have

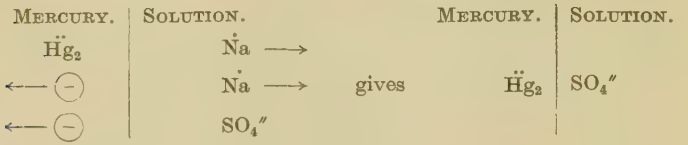
$$\frac{d\gamma}{d\psi} = n\mathrm{F}\Gamma_{\dot{\mathrm{H}}\dot{\mathrm{g}}_2} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (2)$$

where n equals 2 and is the valency of a mercurous ion and F is the Faraday. This is an equation of the same form as the Lippmann equation if $\mathrm{Q} = n\mathrm{F}\Gamma_{\dot{\mathrm{H}}\dot{\mathrm{g}}_2}$.

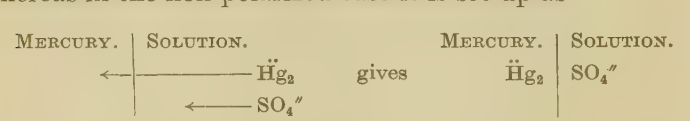
Owing to the reputed failure of the Lippmann-Helmholtz theory for the polarized interphase, this Gibbs type theory has been applied to this case also (for example, see Frumkin [20]). It was considered that when the polarizing potential was applied the concentration of mercurous salt in the layer of solution next to the interphase became adjusted by electrolysis to the equilibrium value for the applied potential, as calculable from the Nernst theory of electrode potential, and that then the Gibbs theory could be applied to give the properties of the interphase. Thus all polarization was considered to be concentration polarization. In the first place this is unnecessary, and, moreover, recent work by Andauer and Lange [21] shows that when the mercurous ion concentration is below 10^{-8} N it ceases to be effective in completely determining the potential of a mercury electrode. And experimental work like that of Rosenberg and Stegeman, and that described in the present paper, shows that for practical purposes the potential of an expanding mercury surface is only determined by the mercurous ion concentration in the solution when this is greater than about N/1000, even for very slow rates of expansion. On the ordinary electrocapillary curves for a polarized interphase the concentration of mercurous ions that would be in equilibrium goes down to 10^{-60} N. Andauer and Lange [22] therefore point out that for a polarized interphase, whether in a dropping electrode experiment or in a capillary electrometer, a Lippmann method is the correct one and a Gibbs method is definitely wrong, even though the result may be correct if suitable definitions are made of the quantities involved.

Thus we are left with the fact that for the restricted

ranges of potential in which curves may be obtained which obey equation (2), they practically coincide with curves obeying equation (1). This means that for any given potential the electrolytic double layer in both cases has similar properties. In the case of the polarized interphase the double layer may be considered to be set up in the way indicated diagrammatically as



whereas in the non-polarized case it is set up as



The presence of other ions in the solution does not affect this argument, because their adsorption from the solution would be much the same for both cases. Thus for both cases the double layer has the same final form and $n\mathrm{F}\Gamma_{\ddot{Hg}_2}=\mathrm{Q}$. In all cases, therefore, the slope of the electrocapillary curve gives Q, the charge on the double layer at the interphase, and this forms a basis on which electrocapillary data may be interpreted in order to elucidate the properties of the double layer itself and to obtain information about other substances that may be adsorbed in it.

Summary.

(1) Electrocapillary curves for solutions containing mercurous salt have been obtained by the drop-weight method. It has been shown that the potential of the dropping mercury differs from that of a stationary mercury electrode in the same solution, and that if the surface tension is plotted against the measured potential of the drop, the curve has no maximum in the region of $\epsilon_{\mathrm{H}}=+0{\cdot}5$ volt. Thus there is no true electrocapillary maximum in this region as had previously been claimed from work in which the surface tension was plotted against the potential of the stationary mercury electrode.

(2) Hence the evidence of two of the most important

methods that were thought to indicate that the potential of the electrolytic null-point (absolute zero of potential) was near $\epsilon_H = +0{\cdot}5$ volt is really without value for that purpose.

(3) On this basis the theory of electrocapillarity is discussed and the Gibbs method of treatment shown to be inapplicable to the case of the polarized interphase, whose properties are best given by the Lippmann treatment.

(4) It has been shown that the electrocapillary curves for a polarized and for a non-polarized interphase are really the same for the limited range in which they can be compared, and this shows that the electrolytic double layer is probably also the same for both cases.

Finally, the author wishes to thank Mr. Oliver Gatty for valuable criticism and advice, and the Department of Scientific and Industrial Research for a grant which enabled this work to be carried out.

References.

(1) Nernst, *Zeit. Elektrochem.* iv. p. 29 (1897).
(2) Bennewitz and Delijannis, *Zeit. Phys. Chem.* cxxv. p. 144 (1927).
(3) Rosenberg and Stegeman, Journ. Amer. Chem. Soc. xxx. p. 1306 (1926).
(4) Frumkin and Obrutschewa, *Zeit. Phys. Chem.* A, cxxxviii. p. 246 (1928).
(5) Bennewitz and Küchler, *Zeit. Phys. Chem.* A, cliii. p. 443 (1931).
(6) Harkins and Brown, Journ. Amer. Chem. Soc. xli. p. 499 (1919).
(7) Kučera, *Ann. Physik*, xi. pp. 529 & 698 (1903).
(8) Desha, Amer. Chem. Journ. xli. p. 152 (1909).
(9) Gouy, *Ann. Chim. et de Phys.* (7) xxix. p. 145 (1903); *Ann. Physique*, (9) vi. p. 5 (1916).
(10) Stern, *Zeit. Elektrochem.* xxx. p. 508 (1924).
(11) Guggenheim, Journ. Phys. Chem. xxxiii. p. 842 (1929); xxxiv. p. 1540 (1930).
(12) Koenig, *Zeit. Phys. Chem.* A, cliv. pp. 421 & 454 (1931).
(13) Philpot, Phil. Mag. (7) xiii. p. 775 (1932).
(14) Paschen, *Ann. Physik*, xliii. p. 585 (1891).
(15) Bennewitz and Schulz, *Zeit. Phys. Chem.* cxxiv. p. 115 (1926).
(16) Bennewitz and Bigalke, *Zeit. Phys. Chem.* A, cliv. p. 113 (1931).
(17) Billiter, *Zeit. Phys. Chem.* xlviii. p. 513 (1904).
(18) Lippmann, *Ann. Chim. et de Phys.* (5) v. p. 494 (1875).
(19) Helmholtz, *Gesamm. Abh.* i. p. 925 (1879).
(20) Frumkin, *Ergebn. d. exakt. Naturw.* vii. p. 240 (1928).
(21) Andauer and Lange, *Zeit. Phys. Chem.* A, Bodenstein Festband, p. 251 (1931).
(22) Andauer and Lange, *Zeit. Phys. Chem.* A, clxii. p. 241 (1932).
(23) Craxford, Gatty, and Philpot (to be published shortly).
(24) Frumkin, *Zeit. Phys. Chem.* ciii. p. 43 (1922).

Balliol and Trinity College Laboratory,
Oxford.

VI. *The Influence of Surface Energy in Disperse Systems.* *By* Chu-Phay Yap †.

SINCE Bakhuis Roozeboom [1] showed the practical applications of the phase rule, as deduced by Gibbs [2] by means of the concept of "chemical potential," the fundamental importance of the principle as a guide to the understanding of the nature of the equilibria in heterogeneous systems has been fully attested by the numerous studies in equilibrium diagrams (hereafter referred to as phase diagrams). Cases have been found in which the phase rule appears to be inapplicable, and it is, therefore, pertinent at the outset to keep in mind certain facts which are explicitly assumed by Gibbs in deriving the rule of phase, namely :—

(1) No phase must be present in such a fine state of subdivision that its fugacity is appreciably influenced by surface forces (*e. g.*, surface tension).

(2) It is assumed that the only variables are pressure, temperature, and composition, thus excluding the influence of electrostatic, electromagnetic, and gravitational fields. Moreover, it is assumed that all phases in the system are under the same pressure throughout.

It is the main purpose of this paper to treat from strictly formal thermodynamic point of view the effect of subdivision on the equilibria of heterogeneous systems and to correlate the various influences of surface energy in disperse systems.

Difference between Equilibrium in Aggregated and in Disperse Systems.

Let us assume a saturated solution of A and B (designated hereafter by the subscripts 1 and 2 for convenience) in contact with some excess solid A in the aggregated form. If we denote, following the conventions established by Lewis and Randall[3], by F_1' the molal free energy of the solid and by $\bar{F}_1'$ its partial molal free energy in the solution, then since the system is in equilibrium, $F_1'=\bar{F}_1$ and so $dF_1'=d\bar{F}_1$. We may write for the solution

$$d\bar{F}_1=\frac{\partial\bar{F}_1}{\partial T}dT+\frac{\partial\bar{F}_1}{\partial P}dP+\frac{\partial\bar{F}_1}{\partial N_1}dN_1, \quad . \quad . \quad . \quad (1)$$

† Communicated by the Author.

while for the pure solid phase, likewise

$$d\mathrm{F}_1' = \frac{\partial \mathrm{F}_1'}{\partial \mathrm{T}} d\mathrm{T} + \frac{\partial \bar{\mathrm{F}}_1}{\partial \mathrm{P}} d\mathrm{P}. \quad . \quad . \quad . \quad . \quad (2)$$

We note that both the partial molal free energy of A in the solution and its partial molal free energy in the solid state are defined by the three variables—pressure, temperature, and composition. As the composition is calculated on the basis of mol-fraction or mol per cent., we see above that in a two-component system, when the mol-fraction of A is known, the mol-fraction of B is fixed as $(1-\mathrm{N}_1)$; likewise, in a three-component system, the mol-fraction of only two molecular species need be known. Hence, for any phase, the number of variables is $(n-1)+\mathrm{P}+\mathrm{T}$, and for p number of phases, $p(n-1)+2$. In the same way, since $d\mathrm{F}_1'=d\bar{\mathrm{F}}_1$ for a two-phase equilibrium, we need, therefore, know only one independent free energy equation, as the other is always equal to it; hence, for n number of components we need to know $n(p-1)$ free energy equations. Since the degree of freedom is defined as the difference between the number of variables and the number of independent free energy equations therefore,

$$f=[p(n-1)+2]-n(p-1)$$

or

$$=n-p+2, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (3)$$

which is Gibbs's phase rule.

Let us now suppose that the solid A is subdivided into extremely small particles, so that the influence of surface forces becomes appreciable. Then for the solution we have, as before,

$$d\bar{\mathrm{F}}_1 = \frac{\partial \bar{\mathrm{F}}_1}{\partial \mathrm{T}} d\mathrm{T} + \frac{\partial \bar{\mathrm{F}}_1}{\partial \mathrm{P}} d\mathrm{P} + \frac{\partial \bar{\mathrm{F}}_1}{\partial \mathrm{N}_1} d\mathrm{N}_1^*, \quad . \quad . \quad (4)$$

while for the pure subdivided solid phase

$$d\mathrm{F}_1' = \frac{\partial \mathrm{F}_1'}{\partial \mathrm{T}} d\mathrm{T} + \frac{\partial \mathrm{F}_1'}{\partial \mathrm{P}} d\mathrm{P} + \frac{\partial \mathrm{F}_1'}{\partial \sigma_1} d\sigma, \quad . \quad . \quad . \quad (5)$$

in which we have now introduced an additional variable, σ, for the surface. Hence, for any phase, the number of variables is now $[(n-1)+\mathrm{P}, \mathrm{T}, \sigma]$, and so the usual phase rule should then become

$$f=n-p+3, \quad . \quad . \quad . \quad . \quad . \quad (6)$$

which is equivalent to saying that a disperse system is unstable towards the aggregated system and that the former is univariant with respect to the latter.

Tolman [4] has derived a generalized phase rule for aggregated and disperse systems, in which the degrees of freedom

$$f=n-r+v, \quad . \quad . \quad . \quad . \quad . \quad . \quad (7)$$

where v is the smallest number of variables, in addition to the composition, which are necessary to define the system, and r is the number of regions. A region (throughout which the properties of matter are either uniform or are a continuous function of the macroscopic position) is said to be divided from the rest of the system by boundary surfaces where an abrupt change in the properties can occur. Suppose we have two particles of subdivided A with surface σ' and σ'', then we have two regions, and the system can only be equilibrated if they become identical in surface, or have at least three variables different. Tolman comes to the conclusion that the addition of each unduplicated region † introduces an additional degree of freedom and that for equilibrium all particles must be identical. Hence, for an isodisperse system the ordinary phase rule of Gibbs applies, and the thermodynamic distinction between (1) a colloidal and a true solution and (2) a colloidal and aggregated state vanishes. All these are implicit in equation (5), because when σ is constant the last differential coefficient drops out. Moreover, equation (6) could have been developed similarly in a generalized way to give equation (7).

By combining equations (1) and (2) and substituting the proper values of the differential coefficients, we obtain at constant pressure

$$d\ln N_1=\frac{\Delta H}{RT^2}dT, \quad . \quad . \quad . \quad . \quad . \quad (8)$$

which is the Raoult-van't Hoff law of the depression of freezing-point. Likewise, upon combining equations (4)

† It is not quite clear whether Tolman uses the term unduplicated in the sense of, say, difference in σ, or in the sense of non-identity of the influences which affect the equilibrium of the system. If used in the former sense, then it is incorrect, as the mere presence of only two subdivided particles of unequal size introduces an additional degree of freedom, and the further introduction of particles of varying sizes does not affect the degrees of freedom in the least.

and (5), keeping in mind that $(\partial F/\partial\sigma)=\gamma$, we obtain, again at constant pressure,

$$d\ln N_1^*=\frac{\Delta Q}{RT^2}dT+\frac{\gamma_1 d\sigma}{RT}, \quad . \quad . \quad . \quad . \quad (9)$$

in which ΔQ is the heat of solution of the finely divided A and N_1^* is the super-solubility. If we determine the normal solubility and the supersolubility of A at two temperatures not far apart, T_1 and T_2, keeping σ constant, then upon integrating equations (8) and (9) and subtracting so that

$$q=\Delta H-\Delta Q=\frac{(\ln N_2/N_1)-(\ln N_2^*/N_1^*)}{R(1/T_1)-R(1/T_2)}, \quad . \quad (10)$$

in which q is the heat equivalent of the formation of the surface σ. Knowing the particle-size, therefore the value of σ, we can thus obtain the total surface energy

$$u=q/\sigma=\gamma-T(d\gamma/dT).$$

The value of q can also be obtained calorimetrically by determining the difference between the normal heat of solution and the heat of solution of the subdivided solid at constant temperature, as was done by Lipsett, Johnson, and Maass [5]. The direct determinations of the normal solubility and the supersolubility at two different temperatures are not only much easier to carry out, but also furnish us with an indirect check of the value of u from the value of q/σ and from $\gamma-T(d\gamma/dT)$.

If we are interested in the change in solubility with particle-size at constant temperature, then the first term on the right-hand side of equation (9) drops out and integrating between N_1, the normal solubility when the σ is small enough to be considered zero and the super-solubility N_1^*, when the surface is equal to σ^*, we obtain

$$\left.\begin{aligned}\int_{N_1}^{N_1^*} d\ln N_1&=\frac{\gamma}{RT}\int_0^{\sigma^*} d\sigma,\\ RT\ln(N_1^*/N_1)&=\gamma\sigma^*.\end{aligned}\right\} \quad . \quad . \quad . \quad . \quad (11)$$

Equation (11) can also be shown in a very simple manner as follows: let f_1' and f_1 represent respectively the fugacity of the pure solid A in any state of subdivision and the fugacity of A in solution; then, according to our condition of equilibrium, $f_1'=f_1$ and likewise $d\ln f_1'=d\ln f_1$. At constant temperature and pressure,

the fugacity of the pure solid A depends only upon the state of subdivision, while its fugacity in solution depends only upon its concentration N_1; hence

$$\frac{\partial \ln f_1'}{\partial \sigma} d\sigma = \frac{\partial \ln f_1}{\partial N_1} dN_1. \quad . \quad . \quad . \quad . \quad . \quad (12)$$

As $(\partial \ln f_1'/\partial\sigma)=\gamma/RT$ and $(\partial \ln f_1/\partial N_1)=1/N_1$, equation (12) reduces to

$$\frac{\partial N}{N_1} = \frac{\gamma}{RT} d\sigma, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (13)$$

which is identical with equation (11), in which pressure and temperature are understood also to be constant.

The change in the free energy (which is equivalent to the work done) of a molal volume of A upon subdividing it to fine particles is evidently equal to the surface tension times the change in the surface with respect to the volume of the particles, that is,

$$F-F^\circ=\gamma(d\sigma/dv)V. \quad . \quad . \quad . \quad . \quad . \quad . \quad (14)$$

But $(d\sigma/dv)=2/r$, and as the molal free energies of A in the two states are equal to the partial molal free energies in the solutions of N_1^* and N_1, that is,

$$F-F^\circ=RT \ln (N_1^*-N_1).$$

Substituting this value in equation (14) gives

$$RT \ln (N_1^*-N_1)=2V\gamma/r, \quad . \quad . \quad . \quad . \quad (15)$$

which is the well-known Kelvin-Freundlich [6] equation relating supersolubility to the particle-size of finely subdivided solids. It should be noted that in the treatment above we have considered, as usual, that the surface of the aggregated solid is zero.

In case the solid is a compound which upon dissolution dissociates to the extent of α, giving n number of ions, we have the Dundon and Mack [7] equation that

$$(1-\alpha+n\alpha)RT \ln (N_1^*-N_1)=2V\gamma/r. \quad . \quad . \quad (16)$$

The Phenomenon of Undercooling.

We may define undercooling as a case of delayed precipitation and consequent supersaturation, brought about by the influence of particle-size of the first precipitate formed. Since the particle-size of the nuclei (precipitate) is in the range in which surface forces are

operative, it is, therefore, important to ascertain the change of particle-size with respect to the temperature to which a solution may be undercooled. Suppose we have a solution of a definite composition N_1, which we allow to undercool without precipitation, so that the surface of the particles of A which can be in equilibrium with the solution according to equation (15) changes with the temperature. Then

$$\left(\frac{\partial\sigma}{\partial T}\right)_N = -\frac{(\partial N/\partial T)_\sigma}{(\partial N/\partial\sigma)_T} = -\frac{(\partial \ln f/\partial T)_\sigma}{(\partial \ln f/\partial\sigma)_T} = -\frac{\triangle H/RT^2}{\gamma/RT} = -\frac{\triangle H}{\gamma T^\circ}, \quad . \; . \; , \quad (17)$$

in which T° is a constant and is the normal temperature at which a solution of N_1 concentration can be in equilibrium with solid A in the normal state of aggregation, and likewise ΔH is the normal heat of solution. Since $\sigma = 2V/r$, therefore, $d\sigma = -2V dr/r^2$. Removing the restriction of the partial derivative in equation (17) as it is understood we are dealing with constant composition and substituting $-2V dr/r^2$ in place of $d\sigma$, gives

$$\frac{dr}{dT} = \frac{\Delta H r^2}{2V\gamma T^\circ} = kr^2 \quad . \; . \; . \; . \; . \quad (18)$$

or

$$-1/r = kT + C. \quad . \; . \; . \; . \; . \quad (19)$$

At normal temperature T°, $r = \infty$, and the integration constant C is then equal to $-kT$. Substituting it in equation (19) and transposing gives

$$T = T^\circ - 1/kr = T^\circ\left(1 - \frac{2V\gamma}{\Delta H r}\right), \quad . \; . \; . \quad (20)$$

which in effect is identical with an equation developed by Jones and Partington [8] by a different method. According to equation (20), whether $T <> T^\circ$ depends entirely on the sign of the heat of solution, while the degree of undercooling depends on the magnitude of $(2V\gamma/\Delta H r)$. As generally the heat of solution is positive (involving absorption of heat) $T < T^\circ$, that is, we have undercooling, which may be pictured as follows : If the radius of the nuclei is less than the equilibrium radius r, the temperature decreases without precipitation and the solution consequently becomes supersaturated. In such solutions the tendency towards crystallization is, however, extremely small at a low degree of supersaturation. The

smaller nuclei ($<r$) which could form comparatively easily due to molecular collisions, are unstable, however, and dissolve again immediately; the larger particles ($>r$), which are stable, are not readily formed on account of their size, even in spite of the probably high velocity of crystal growth. Upon further undercooling, the equilibrium particle size r decreases similarly † and the solution becomes even more supersaturated, so that the conditions become highly favourable for final crystallization and precipitation.

Stability of Disperse States.

A system is said to be in equilibrium with respect to every possible change when the free energy remains unchanged in any infinitesimal process occurring at constant pressure and temperature, that is, $d\mathrm{F}=0$. In different language, when the free energy of the system is a minimum, it is unable to do any spontaneous work; hence, it is in an equilibrium state. Let us now examine under what conditions a disperse state may be considered stable. Following Rice [9] the three kinds of changes possible in ordinary disperse systems are:

(1) *Drop-solution changes.*—Transfer of material from the colloid particles into the dispersing medium and *vice versa*, without change in the number of particles.

(2) *Drop-to-drop changes.*—Interchange of material between particles of different sizes, without change in the number of particles.

(3) *Drop-density changes.*—Interchange of material between the particles and the dispersing medium, or between the particles themselves, involving change in the number of particles.

In our treatment of dispersoids we have considered the surface tension to be constant and independent of the curvature, which is true only when r is fairly large, and that the concentration of the dispersoid per unit volume does not affect its fugacity in the dispersing medium (*i. e.*, we are treating the ratio of the components as an

† Applying the Le Chatelier-Braun principle of mobile equilibrium, we should expect the equilibrium particle-size to decrease with the temperature, as the heat of solution decreases with the particle size and the system is cooling down.

independent variable), which is true when we are not dealing with extremely concentrated colloidal solutions. Besides these considerations, the influence of drop-density change on the free energy of the system may be safely disregarded, as it is essentially only a second-order differential. Rice insists that a disperse system can only be permanently stable when it satisfies the criterion of stability with respect to the drop-density change †.

We have already seen that when the temperature and pressure are defined, an aggregated system is in equilibrium, but in the case of a disperse system another variable must be defined. Therefore, at constant temperature and pressure, a disperse system can only be in equilibrium when the last term in equation (5) drops out, that is, $(\partial F/\partial\sigma)=0$. It is obvious that $(\partial F/\partial\sigma)$ can only be zero when (1) the particles are all of the same size, and (2) the surface tension is zero. In the latter case, as we shall show later, an anisodisperse system can be in equilibrium. A further criterion of equilibrium of an isodisperse system with change in temperature is that $d\mathrm{F}/d\mathrm{T}=0$, where F is the free energy content of a particle of a definite surface (which is constant) and, likewise, when the system is incompressible, $d\mathrm{F}/d\mathrm{P}=0$. When these conditions are fulfilled, an isodisperse system will behave towards temperature and pressure, as if it were an aggregated system, that is, as any other stable system ‡. We shall, therefore, be able to construct a series of the usual phase diagrams for each value of σ, therefore of particle-size. The solubility curve will thus generate a surface in a T–c–σ (c being the composition) space

† It appears that Rice confuses F, the free energy as Lewis uses it, with F, the free energy as Helmholtz formulated it, which Lewis designates as A and is related to the first F by the equation, $F=A-PV$. The drop-density change, according to Rice, involves the work function PV, and what he is dealing with is the change in A and not F of the system.

‡ An objection might be raised on the ground that when a particle accidentally becomes larger than the rest, an isodisperse system is no longer stable. While conceding the force of the argument it should be remembered that everything is in a state of equilibrium only relatively. To give a homely illustration: a stone lying in a depression on the top of a hill is in a state of rest, that is, in equilibrium; yet when someone accidentally kicks it, it will roll downhill until it reaches another depression. On the other hand, the deliberate introduction of a larger particle into an isodisperse system in equilibrium is tantamount to making σ variable again. How long an isodisperse system may remain in equilibrium is certainly not a problem of thermodynamics.

diagram, just as it does in the P–T–c space diagram. We shall show the relations between the two cases later in this paper.

As stated above, when $\gamma=0$ an isodisperse system can be in equilibrium, because in this case the free energy of the subdivided particles will be independent of the particle-size. Let us, therefore, examine under what conditions γ can be reduced to zero. Firstly, we have the case of the formation of a colloidal state (of liquid in liquid) at the critical temperature, as evidenced by the opalescent appearance of the liquid, as Tolman has pointed out. Secondly, we may be able to form such a colloidal state in the case of two normally incompletely soluble liquids at the temperature when they are just completely miscible in each other. Thirdly, we may have the case of the Helmholtz double layer effect of an electrical nature due to adsorption, which may give rise not only to zero surface tension, but also presumably to negative surface tension †. As Knapp [10] has shown, the relation between particle-size and solubility, when the particles carry a (assumed as rigid) double layer of h thickness, K dielectric capacity, and carrying q charge, is

$$\frac{RT}{V}\ln(N_1{}^*-N_1)=\frac{2\gamma}{r}-\frac{q^2h}{4\pi K r^5}. \quad . \quad . \quad (21)$$

We note that the left-hand side of equation (21) is $k(RT/V)$, where k has no dimensions, and so the two terms on the right-hand side must have the dimensions of pressure to be correct, which was found to be so. Hence, we designate $2\gamma/r$ as p_γ, the pressure due to the atomic surface configuration, and $q^2h/4\pi K r^5$ as p_ϵ, the pressure of an electrical nature. These two pressures are opposed to each other, and the resulting surface pressure will be zero when

$$2\gamma/r=q^2h/4\pi K r^5, \quad . \quad . \quad . \quad . \quad . \quad (22)$$

that is, when the contractile force due to p_γ is just counterbalanced by the expansive pressure due to p_ϵ. Intro-

† Donnan [Phil. Mag. i. p. 647 (1901)] has also advanced arguments regarding the probable influence of negative surface tension, as has Tolman (*loc. cit.*), who claims that those substances which peptize must possess a negative surface tension with respect to the dispersing medium. While the concept of a negative surface tension is mathematically interesting, it is doubtful if it has any real physical significance.

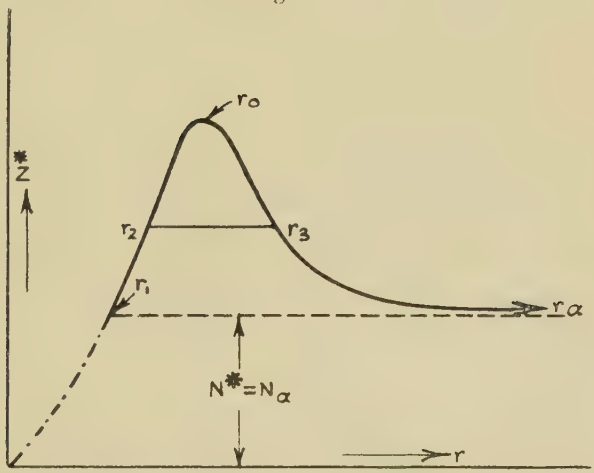
ducing the charge density of the surface, $\rho = q/4\pi r^2$ in equation (22) gives

$$\rho^2 = \frac{K\gamma}{2\pi h}, \quad . \quad . \quad . \quad . \quad . \quad . \quad (23)$$

an equation which Lewis [11] derives to prove that a critical particle-size is not required. This is, however, obvious from our previous consideration for the stability of a polydisperse system †. What is not quite obvious is that from equation (21) when we set ($N_1{}^* = N_1$) so that

$$r_1 = +\sqrt[4]{\frac{q^2 h}{8\pi K\gamma}}, \quad . \quad . \quad . \quad . \quad . \quad (24)$$

Fig. 1.



The graph of equation (21). When $p_\gamma = p_\epsilon$, then $N^* = N_\infty$. Only particles with radii r_1 to r_∞ have any real physical significance.

which has been erroneously interpreted as the single critical value of the particles which can be in equilibrium with the aggregated solid. What it simply means is that when the upper limit of particle-size is placed at $r = \infty$, the lower limit of particle-size is given by equation (24), as shown graphically in fig. 1. If the upper limit of

† In fact, any other equation which does not satisfy the criterion that the free energy of the particles is independent of their size when the resultant surface pressure is zero, cannot be correct. Thus Knapp (*loc. cit.*) and Lewis [*Z. Physik. Chem.* xxxiv. p. 495 (1900)], in treating the particles as carrying a single electrically charged and insulated layer, derived the value of $p_\epsilon = q^2/8Kr^4$, which obviously cannot satisfy the above criterion.

particle-size were placed at r_3, corresponding to N_3 in concentration, then the lower limit is correspondingly r_2 when $N_2=N_3$, as shown in fig. 1, the range of stable particle-size decreases with increasing concentration, so that at the maximum solubility only one particle-size r_0 is stable. This is the most stable state of dispersion for a system of this type.

It is noteworthy that from fig. 1 particles of radius less than r_1 are either fictitious, without any real physical significance, or they possess a negative surface tension, that is, $p_\kappa < p_\gamma$; hence, they should go into solution, which is thus unsaturated. This is in distinct contradiction to one of Tolman's conclusions regarding the influence of negative surface tension. Another interesting implication of fig. 1 is that with increasing radius of the particles supersolubility decreases very rapidly. This was apparently confirmed by the experiments of Jones and Partington (*loc. cit.*), who found no supersolubility with particles larger than 50 μ. It is also possible to infer from equation (21) and fig. 1 that by the Bredig method of preparation of metallic colloidal sols, we might expect the particle's size to vary from r_1 to $r\infty$, as it is well known that even in such a "neutral" dispersing medium as water selective adsorption takes place giving rise to the electrical double layer. Moreover, it is probable that the particle-size distribution will conform to some kind of probability curve.

It does not appear necessary to postulate with Tolman the presence of a negative tension in those substances which peptize in water, *e. g.*, gelatin and dextrin. In the process of dissolution the subdivided particles which are dissociated from the aggregated mass immediately acquire a Helmholtz double layer, and, as we see from equation (23), we should expect the particles to readjust their size to the charge on it, as Lewis has properly pointed out. This would be consistent with our general conception of the origin of the charge as a result of selective adsorption on the surface of the particles.

An Isodisperse State viewed as a System under Negative Pressure.

It can be easily shown, as in the development of equation (13) regarding the change in solubility with respect to the

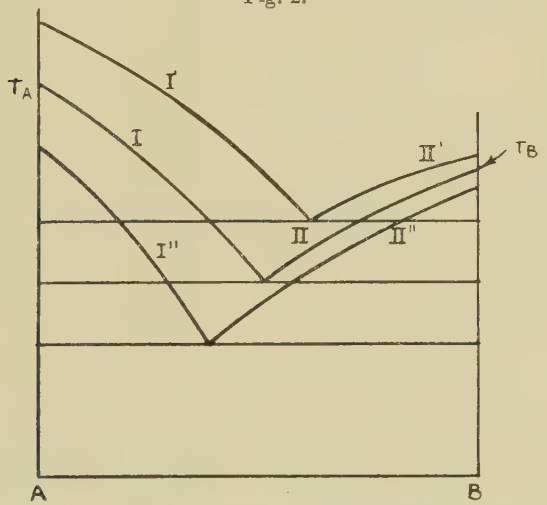
surface, that the change in solubility with pressure is likewise given by

$$\left(\frac{\partial \ln N_1}{\partial P}\right)_{P,T} = \frac{V_1 - V_1}{RT}, \qquad \left(\frac{\partial \ln N_2}{\partial P}\right)_{P,T} = \frac{V_2 - V_2}{RT},$$

. . . (25) and (26)

in which $(V - \bar{V})$ is the difference between the molal volume of solid A or B and its partial molal volume in solution, and as the latter is generally larger than the former the solubility of both A and B should decrease (not equally, however) with increasing pressure.

Fig. 2.



The phase diagram of the system A–B under ordinary, high positive and negative pressures. The phase diagram I″.–II″. also represents an isodisperse system of a definite value of σ.

Treating the composition number c as $(n-1)$, a binary system can be represented in a three-dimensional space diagram of P–T–c components; hence, an ordinary phase diagram of a binary system is merely a plane section of the space diagram at P=1. When we have the additional variable of surface (σ), a three dimensional diagram may also be constructed similarly by treating one variable as constant, *e. g.*, we may construct a T–c–σ space diagram at constant pressure, say, at 1 atmosphere.

In fig. 2, we have the ordinary phase diagram shown

by the solubility curves I. and II., intersecting at the eutectic temperature and composition. Assuming A and B to be normal, melting with an increase in volume, the effect of pressure at constant temperature is, as stated above, to shift the solubility curves upwards, and we should expect in general the composition of the eutectic to shift in the direction of the component whose melting temperature is increased by pressure by the smaller amount, as shown by curves I′. and II′. This was exactly the case as investigated by Pushing and Grevenshikov [12] of the system urethan-diphenylamine ; except from the slope of the solubility curves at high pressures, one concludes that the heat of solution (practically equivalent to the heat of fusion of the pure components) first decreases with the pressure and then practically remains constant. According to Bridgman [13], however, in a large majority of cases the latent heat rises with the pressure, the increase being more pronounced at first, then more gradually with increasing pressure. The curves I′. and II′. are, therefore, drawn on the assumption that the heat of solution increases with the pressure, although no material difference is made on the converse assumption.

If we now extend the space diagrams backwards to a finite negative pressure of $-P$, we will obtain at some value of $-P$ the solubility curves I″. and II″. As should be expected, the solubility curves are depressed unequally as before, and likewise the eutectic composition is shifted towards the A-rich side. The three phase diagrams are superimposed on one another in order to make the relations in the space diagram clearer.

When we construct a T–c–σ space diagram, the normal diagram will be given by I. and II. as before, when σ is set equal to zero as $r=\infty$. At some other value of σ, the phase diagram will be that given by curves I″. and II″. It is implicit in the system A—B that the components have quite different internal pressure, and consequently differ in molal volume, hardness, compressibility, surface tension, etc. At constant surface the last term in equation (5) and in equation (9) drop out; and accordingly, from equation (9) the change in solubility with temperature is a function not of ΔH, the normal heat of solution (fusion), but of ΔQ, the heat of solution of the subdivided solid. As the difference between

ΔH and ΔQ is q, which is a function of u, therefore of γ, we should then expect the solubility curve of the component whose surface tension is larger—A in this case—to be shifted more, and to show a larger change in the slope.

As the solubility curves I''. and II''. can be obtained either by applying a negative pressure, or by subdividing the solids with which the solutions are in equilibrium, a disperse state, therefore, behaves as if it were under a negative pressure, which implies that the phase occupying the larger volume is under such circumstances the more stable one. Due to the influence of surface force, the process of subdivision should be accompanied by a decrease in volume, that is, an increase in density ; hence, we should expect a disperse state to be unstable with respect to the aggregated state, which is in accord with our previous discussion †.

Some Practical Applications of the Concept of Surface Energy.

Inasmuch as the writer's interest is mainly in metallic systems, it may be worth while to call attention to some applications of the concept of surface energy to the problems of equilibria in metallic systems, as this field has been sadly neglected. Thus, for example, in the case of steel, the iron carbide (Fe_3C) may be precipitated from solid solution in all stages of dispersion, from practically submicroscopic particles in β-martensite to well-formed lamellar crystals in the pearlite eutectoid. By suitable heat-treatment, the colloidal carbide in the former may eventually be coagulated to fairly large spherulites. Benedicks [14] apparently was the first one to recognize the colloidal nature of the dispersion of Fe_3C in troostite alone, although Alexander [15] clearly recognized many phenomena in metals and alloys as of a truly colloidal nature.

† Lewis (*loc. cit.*) also comes to the same conclusion regarding the presence of a negative pressure inside a particle carrying a double layer. Some of his assumptions appear to be questionable, as the surface pressure is, according to equation (22), explicitly set at zero in order to show that the particles under such circumstances are polydispersed. The treatment given above applies only to particles without a double layer, that is, for positive values of p_γ, or at least p_γ should be larger than p_ϵ, when a double layer is present. When we have a positive p_γ at the surface, which is directed inwards, at equilibrium, it must be counteracted by an equal pressure directed outwards, in order that during the process of subdivision the molal volume of the solid is not decreased.

A large class of alloys owe their industrial importance to the fact that they are susceptible to precipitation hardening, *i. e.*, the supersaturated alloy quenched from a high temperature is allowed to age at low temperatures to produce desirable hardness, tensile strength, etc. Since hardness and surface energy are both related to internal pressure, we should expect the metals and their compounds to possess high surface energies, particularly the latter. This fact forms the basis of the writer's theory [16] that the hardness of precipitation alloys may be attributed mainly to the surface energy latent in the highly dispersed precipitate †.

One of the fundamental problems of physical metallurgy has been the exact determination of the solid solubility curves. For example, the solid solubility of Ag in Cu has been determined by numerous investigators, yet hardly two workers can fully agree in their results in spite of extreme refinements in their experimental technique and prolonged annealing to obtain equilibrium. In such studies the precipitates (assumed to be in equilibrium with the solid solution) are generally in such a state of subdivision that the surface forces are operative. According to equation (15) we note that, depending upon the particle-size, the solid solution may be supersaturated to different degrees, and therefore the influence of surface energy has been entirely overlooked, as the writer has already called attention to it elsewhere [17].

In many studies in equilibrium between gases and solids, it will be readily apparent that the equilibrium constants obtained may not represent the true equilibrium of the reactants and the resultants in their standard states‡, as shown by the following considerations: let F° be the standard free energy of the reaction, F' the free energy obtained when the solid is in a state of subdivision, and f the free energy due to the surface of the subdivided solid. Then

$$F^\circ = F' - f = -RT \ln K' - f, \quad . \quad . \quad (27)$$

† The prevailing theories of hardness of such alloys emphasize unduly a mechanistic picture of slip-interference and lattice distortion, as these phenomena are easy to visualize and present in simple engineering terms of strain, friction, etc. The rôle of surface energy, while very difficult to visualize, is an attempt to explain the hardness of metallic disperse states by taking into account the unbalanced forces at the interface.

‡ For a definition of standard states of gases, liquids, and solids, consult 'Thermodynamics,' by Lewis and Randall, p. 291.

that is, the free energy obtained will always be larger than that referred to the standard state by F. Thus, in the reaction C (graphite)$+CO_2 \rightleftharpoons 2CO$, the backward reaction may result in extremely fine graphite being produced, so that the final equilibrium is really between the colloidal graphite and the carbon oxides. Consequently, the free energy of the reaction thus obtained will be larger than that with aggregated graphite, and as free energy equations, like the thermochemical equations, can be added and subtracted, an error is thus introduced in free energy calculations without provision for any form of correction.

In the writer's study on the free energy of formation of Fe_3C [18] he found this difficulty. Thus, in the reaction

$$3(\alpha)Fe+C\ (graphite)=Fe_3C,$$

it is understood that all these substances are in their standard state; yet, actually, the surface energy not only of the graphite, but also of Fe_3C, may be involved, as in

$$3(\alpha)Fe+2CO=Fe_3C+CO_2,$$

the iron carbide in the steel may range in size down to where the surface energy becomes appreciable. Hence, the standard molal free energy of the formation of Fe_3C may be smaller than the observed value by the difference in F(C) and F(Fe_3C).

Summary.

1. An extension of Gibbs's phase rule to take into account the influence of surface has been presented in some detail. A system in which one of the phases is in a state of subdivision (particle-size not being constant, however) has an additional degree of freedom, that is, pressure and temperature are not enough to define the composition of the system; hence, $f=n-p+3$.

2. Equations relating solubility, heat of solution, etc., to particle-size have been given.

3. The phenomenon of undercooling has been discussed from the standpoint of equilibrium particle-size and of the change in the equilibrium particle-size, therefore, surface, with the temperature. An equation, somewhat similar to one developed by Jones and Partington relating the change in the equilibrium particle-size with temperature of undercooling, has been developed.

4. The stability of disperse states has been discussed from the point of view of an isodisperse state and of an anisodisperse state when the surface tension is zero. In the case of particles with a Helmholtz electrical double layer the surface force is equal to zero when the surface tension is just counterbalanced by the electrical surface force, that is, $p_\gamma = p_\epsilon$.

5. A disperse state behaves as if it were under a negative pressure in opposition to the positive pressure due to surface tension (directed inwards). This implies that the state occupying the larger volume is the more stable one, that is $(\partial F/\partial(-P)) = -V$, and as a disperse state is unstable with respect to the aggregated state, the process of subdivision would have been accompanied by a decrease in volume and an increase in density were it not for the fact that the contractile tendency of the surface tension is counter-balanced by the negative pressure.

Some practical applications of the influence of surface energy to the problems of metallic systems have been shown.

References.

(1) 'Die heterogenen Gleichgewichte, Braunschweig,' 1901 *et seq.* (Original not read.)
(2) 'The Collected Works of J. Willard Gibbs' (Longmans, Green & Co.: New York, 1926), i. p. 55 *et seq.*
(3) Lewis & Randall, 'Thermodynamics' (McGraw-Hill Book Co. Inc.: New York, 1923).
(4) J. Amer. Chem. Soc. xxxv. pp. 306, 317 (1913).
(5) *Ibid.* xlix. p. 925 (1927).
(6) Freundlich, 'Colloid & Capilliary Chemistry.' Translated by H. S. Hatfield. (Methuen & Co. Ltd: London, 1926, p. 154.)
(7) J. Amer. Chem. Soc. xlv. p. 2479 (1923).
(8) Phil. Mag. xxix. p. 35 (1915).
(9) J. Phys. Chem. xxx. p. 189 (1926).
(10) Trans. Faraday Soc. xvii. p. 457 (1921).
(11) *Ibid.* xxviii. p. 597 (1932).
(12) *Z. Physik. Chem.* cxviii. p. 276 (1926).
(13) Bridgman, 'The Physics of High Pressure' (The MacMillan Co.: New York, 1931), p. 201.
(14) *Z. Physik Chem.* lii. p. 733 (1905).
(15) For a *resume* of Alexander's views regarding colloidal phenomena metals and alloys, consult 'Colloid Chemistry, Theoretical and Applied,' edited by him. (The Chemical Catalog Co. Inc.: New York, 1931, iii. p. 413.)
(16) *Ibid.* p. 463. Also Trans. Amer. Soc. Steel Treat. xx. p. 289 (1932).
(17) Trans. Amer. Inst. Min. Met. Engrs., Inst. Met. Div., 1931 volume, p. 49.
(18) Trans. Faraday Soc. xviii. p. 788 (1932).

Department of Chemistry,
Washington Square College,
New York University,
New York City.

VII. *On the Determination of some of the Elastic Constants of Rochelle Salt by a Dynamical Method.* *By* R. M. Davies, *M.Sc., Lecturer in Physics, University College of Wales, Aberystwyth* *.

Abstract.

The theory of the method employed is given, and various theoretical problems concerning Rochelle Salt piezo-electric resonators are described and discussed. The problems discussed include those relating to the orientation of the resonators with respect to the crystal axes and the correction of the simple frequency formula for finite lateral dimensions, secondary electrical effects, and adiabatic conditions. The correction for finite lateral dimensions is considered in detail, and the result applicable to the general case of a rectangular parallelepiped of crystalline material is given.

The experimental method employed is then described, and the results for 15° C. compared with those calculated from statical experiments on the elastic moduli. The results for the temperature-variation of resonant wave-length show that for two types of resonator this quantity is a linear function of temperature between 0° C. and 20° C. A pronounced discontinuity occurs in the neighbourhood of 23° C., and, in the one case investigated at low temperatures, another discontinuity is found between −14° C. and −15° C. Other physical properties of Rochelle Salt are known to exhibit similar discontinuities at these temperatures. The third type of resonator shows a less well-defined discontinuity near 25° C., whilst the relation between resonant wave-length and temperature is not linear.

§ 1. *Introduction.*

The aim of the work described in this paper was to investigate the suitability of Rochelle Salt as a material for the construction of piezo-electric resonators, and also to study any change in the behaviour of such resonators with change of temperature. A detailed and comprehensive account of the properties and uses of piezo-electric resonators (including a bibliography) is available in a recent book by Vigoureux [1], whilst the general properties of Rochelle Salt are described by Mandell [2].

* Communicated by Prof. Gwilym Owen, M.A., D.Sc.

The experiments described below were carried out in 1926 and 1927[2a], and at that time, as now, quartz was almost universally used for piezo-electric resonators. The main advantages of quartz, in this respect, are its excellent mechanical properties and the small effect of temperature variation on the resonant frequency, the temperature coefficient of the latter being of the order of five parts in a million per degree C. The main advantages of Rochelle Salt are its cheapness and its greater piezo-electric response. Large crystals of Rochelle Salt can be obtained by gradual, controlled cooling of a saturated solution, in which suitable small seed crystals are placed ; piezo-electric resonators can be obtained from the crystals by cutting with a wet string and grinding on fine emery paper.

Since Rochelle Salt resonators could be prepared in this way, it was considered to be of interest to study the effect of temperature on the resonant frequency ; this frequency is a function of the dimensions of the resonator, of its density, and of its elastic moduli. Valasek[3] and Schulwas-Sorokin[3a] have shown that the piezo-electric moduli of Rochelle Salt crystals, cut with edges at 45° to the b and c axes, undergo abrupt changes between −20° C. and −15° C., and again between 20° C. and 25° C. Similar changes occur in the values of other physical constants of the crystal. These changes must be due to some alteration in the structure of the crystal, and since elasticity depends essentially on the structure of the crystal, corresponding changes should occur in the elastic moduli, and, therefore, in the resonant frequency. Such changes were found to occur in the experiments described in this paper, and make Rochelle Salt unsuitable for use in the construction of piezo-resonators.

Similar discontinuities occur in the case of quartz[4], the critical temperature in this case being 575° C. ; using X-ray analysis, these changes have been coordinated by Bragg and Gibbs[5] with changes in the crystal structure.

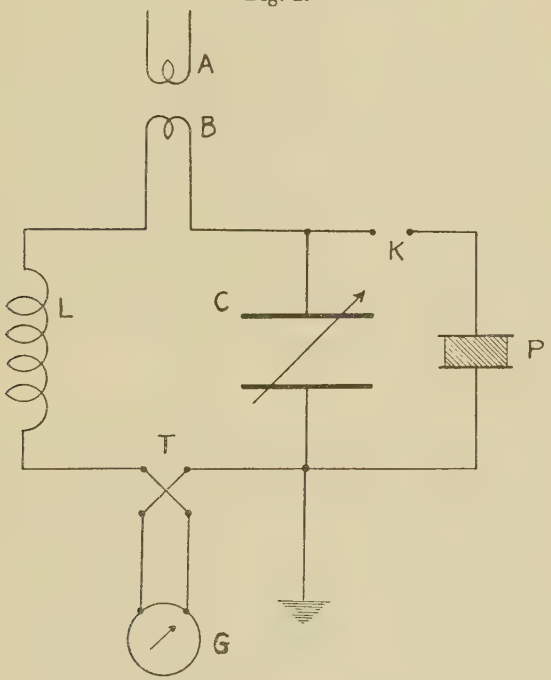
§ 2. *Theory of the Method Employed.*

From a crystal of Rochelle Salt, small rectangular bars are cut so that if a static electric field be applied perpendicular to one pair of faces, linear extension or contraction will occur along the edges bounding these faces. Similarly, an alternating electric field applied in the same direction will give rise to corresponding longitudinal vibrations whose frequency is that of the field. The amplitude of these vibrations will ordinarily be small, since the displacement

per unit field is small; if, however, the frequency of the applied field coincides with one of the mechanical resonant frequencies of the bar, the amplitude becomes very large owing to the small damping factor of the system.

A piezo-resonator consists of a crystal bar cut in this manner and provided with electrodes which are in contact with the appropriate faces, and, which, at the same time, permit free longitudinal vibration. The response frequency

Fig. 1.



Circuit used for the determination of resonant wave-lengths.

of the piezo-resonator can be found by the following method, due to Cady [6].

The plates of the resonator P are arranged so that they can be connected, by the key K, in parallel with the condenser C of a Tinsley wave-meter of the resonant circuit type as shown in fig. 1. The remainder of the wave-meter circuit consists of the main coil L, the vacuo-thermojunction T with its associated galvanometer G, and the inducing coil B. The latter is loosely coupled to a screened valve-oscillator

of variable frequency represented by A, and the current flowing in the circuit is measured by T and G. C is initially adjusted so that the electrical resonant frequency of the wave-meter circuit is approximately that of the piezo-resonator. If, now, the frequency of the valve-oscillator be varied through the response frequency of the piezo-resonator, characteristic changes occur in the current in the circuit; if current be plotted against frequency, the resulting curve shows a narrow crevasse, the bottom of which corresponds to the response frequency of the piezo-resonator. Examples of such curves are given by Cady[7], Dye[8], and Vigoureux[9]. Dye has shown that if the crevasse be present, then the actual frequency-setting of the wave-meter circuit has no effect on the frequency corresponding to the bottom of the crevasse, although it does influence its shape.

The frequency corresponding to the bottom of the crevasse is thus characteristic of the piezo-resonator and not of its associated electrical circuit. This frequency is identical with the mechanical resonant frequency of the bar, as may be seen from the consideration that the crevasse is due to the absorption of energy from the circuit to maintain the vibration of the bar, maximum absorption of energy, *i. e.*, the bottom of the crevasse, occurring at mechanical resonance.

It should be pointed out that there is a slight variation in the response frequency of the piezo-resonator if air-gaps exist between the crystal bar and the electrodes. This point is discussed in detail by Dye[10]; his experimental results show that an increase of 0·1 mm. in the air-gap produces an increase of the order of 1 in 1000 in the response frequency. As described below, this effect was avoided in the present work by arranging that the electrodes made contact with the bar, whilst still permitting longitudinal vibratory motion.

§ 3. *The Orientation of the Bars.*

It was stated above, that, in order that a bar of crystal should be capable of functioning as a piezo-resonator, it is necessary that an electric field applied perpendicular to a pair of faces of the bar should produce linear extension or contraction along the edges bounding these faces. In the case of quartz, it is easy to choose an orientation with respect to the crystal axes which will satisfy the above condition. This is not the case with Rochelle Salt, since an electric field applied parallel to the crystal axes produces a complex system of stresses and strains.

In a Rochelle Salt crystal, let a right-handed rectangular set of axes, Ox, Oy, Oz, be chosen so as to coincide with the $\breve{a}$, $\bar{b}$, c crystallographic axes respectively. If the crystal be subject to an electric field whose components parallel to these axes be $\mathcal{E}_1$, $\mathcal{E}_2$, $\mathcal{E}_3$, then the resulting deformation is given [11] by a set of equations of the type :—

$$\left.\begin{aligned} x_x &= d_{11}\mathcal{E}_1 + d_{21}\mathcal{E}_2 + d_{31}\mathcal{E}_3, \\ &\cdots\cdots\cdots \\ x_y &= d_{16}\mathcal{E}_1 + d_{26}\mathcal{E}_2 + d_{36}\mathcal{E}_3. \end{aligned}\right\} \quad \ldots \quad (3.1)$$

In these equations $x_x \ldots x_y$ are the six components of strain in the usual notation of elasticity ; it may be noted that terms of the type x_x represent longitudinal strains and terms of the type x_y represent shear strains. $d_{11} \ldots d_{36}$ are constants representing strains per unit field, and are termed piezo-electric moduli. Only three of these moduli (d_{14}, d_{25}, d_{36}) differ from zero for crystals of the rhombic hemihedral class to which Rochelle Salt belongs [12].

If we consider the case of an arbitrary right-handed set of rectangular axes Ox', Oy', Oz', not coinciding with the crystallographic axes we may write, with a similar notation,

$$\left.\begin{aligned} x'_x &= d'_{11}\mathcal{E}'_1 + d'_{21}\mathcal{E}'_2 + d'_{31}\mathcal{E}'_3, \\ &\cdots\cdots\cdots \\ x'_y &= d'_{16}\mathcal{E}'_2 + d'_{26}\mathcal{E}'_2 + d'_{36}\mathcal{E}'_3. \end{aligned}\right\} \quad \ldots \quad (3.2)$$

where $x'_x \ldots x'_y$; $\mathcal{E}'_1$, $\mathcal{E}'_2$, $\mathcal{E}'_3$, are the components of strain and of the electric field referred to the arbitrary axes Ox', Oy', Oz', and similarly $d'_{11} \ldots d'_{36}$ are piezo-electric moduli referred to these axes.

To derive the relations between $d'_{11} \ldots d'_{36}$, the piezo-electric moduli $d_{11} \ldots d_{36}$ and the relative orientations of the crystal axes with respect to the arbitrary axes, it is convenient to adopt the following scheme for the direction-cosines concerned :—

	x'	y'	z'
x	α_1	β_1	γ_1
y	α_2	β_2	γ_2
z	α_3	β_3	γ_3

, . . . (3.3)

in which the direction-cosine of the angle between two of the axes is represented by the corresponding Greek letter, *e. g.*, the direction-cosine of the angle between Oz and Oy' is β_3.

The required relations can be obtained as indicated by Voigt[13]; these relations simplify somewhat, in the case of crystals of the rhombic hemihedral class, to the following:—

$$\left.\begin{aligned} d'_{11} &= \alpha_1\alpha_2\alpha_3(d_{14}+d_{25}+d_{36}), \\ &\cdots\cdots\cdots \\ 3d'_{36} &= (2\beta_1^2+\gamma_1^2)(d_{25}-d_{36})+(2\beta_2^2+\gamma_2^2)(d_{36}-d_{14}) \\ &\quad +(2\beta_3^2+\gamma_3^2)(d_{14}-d_{25})+(d_{14}+d_{25}+d_{36})\{\gamma_1(\alpha_2\beta_3+\alpha_3\beta_2) \\ &\quad +\gamma_2(\alpha_1\beta_3+\alpha_3\beta_1)+\gamma_3(\alpha_1\beta_2+\alpha_2\beta_1)\}. \end{aligned}\right\}, \quad \ldots \quad (3.4)$$

Equations (3.2) and (3.4) simplify further if the two coordinate systems have one axis in common, and if the electric field is applied parallel to this axis. Suppose that the Oz and Oz' axes coincide, and that the electric field is of magnitude $\mathcal{E}$ and is applied parallel to this direction. Equations (3.2) reduce to

$$x'_x = d'_{31}\mathcal{E} \ldots x'_y = d'_{36}\mathcal{E}, \quad \ldots \quad (3.5)$$

and we have also that

$$\left.\begin{aligned} &\gamma_1=\gamma_2=\alpha_3=\beta_3=0\ ;\ \gamma_3=1\ ;\ \alpha_1=\beta_2\ ;\ \alpha_2=-\beta_1; \\ &\alpha_1^2+\alpha_2^2=\beta_1^2+\beta_2^2=\alpha_1^2+\beta_1^2=\alpha_2^2+\beta_2^2=1, \end{aligned}\right\} \quad (3.6)$$

with the result that equations (3.4) reduce to

$$\left.\begin{aligned} &d'_{31}=\alpha_1\alpha_2 d_{36}\ ;\ d'_{32}=-\alpha_1\alpha_2 d_{36}\ ;\ d'_{33}=d'_{34}=d'_{35}=0, \\ &d'_{36}=(\alpha_1^2-\alpha_2^2)d_{36}. \end{aligned}\right\} \quad (3.7)$$

Hence the components of strain reduce to

$$\left.\begin{aligned} &x'_x=-y'_y=\alpha_1\alpha_2 d_{36}\mathcal{E}\ ;\ z'_z=y'_z=z'_x=0; \\ &x'_y=(\alpha_1^2-\alpha_2^2)d_{36}\mathcal{E}. \end{aligned}\right\} \quad \ldots \quad (3.8)$$

The shear strain x'_y vanishes if $\alpha_1=\pm\alpha_2$, *i. e.*, if the arbitrary axes Ox' and Oy' make angles of 45° with the $O\breve{a}$ and $O\breve{b}$ crystallographic axes, and the system of strains in this case reduces to longitudinal strains along Ox' and Oy'. It is thus clear that if it be desired to construct piezo-electric resonators in which the electric field is to be applied parallel to the Oc axis, then the edges of the faces perpendicular to this axis must lie at 45° to the $O\breve{a}$ and $O\breve{b}$ axes. For such resonators it can be shown that the *stresses* set up in the crystal by a field $\mathcal{E}$ parallel to Oc reduce to the two tensions X'_x and Y'_y in the direction of the above edges. These tensions are given by the equation

$$X'_x=-Y'_y=-e_{36}\mathcal{E}, \quad \ldots \quad (3.9)$$

where e_{36} is a piezo-electric constant in Voigt's notation. Similar considerations apply to the cases where the applied electric field is to be parallel to the O$\breve{a}$ or O$\bar{b}$ axes; in the former case the edges of the faces perpendicular to the O$\breve{a}$ axis lie at 45° to the Ob and Oc axes; in the latter case, the edges of the faces perpendicular to the O$\bar{b}$ axis lie at 45° to the O$\breve{a}$ and Oc axes.

It is convenient to refer to the resonators as follows :—

Type I.—Applied field parallel to Oc;

Type II.—Applied field parallel to O$\breve{a}$;

Type III.—Applied field parallel to O$\bar{b}$.

§ 4. *The Resonant Frequency of a Piezo-electric Resonator.*

The natural frequencies, n_0, of longitudinal vibration of a thin bar of length L vibrating freely is given by the well-known formula

$$n_0 = \frac{k}{2L}\frac{1}{\sqrt{\rho E}}, \quad . \quad . \quad . \quad . \quad . \quad (4.1)$$

where ρ is the density of the material of the bar, E the extension modulus ("Dehnungsmodul" of Voigt's nomenclature) in the direction of the length of the bar, and k an integer indicating the mode of vibration. E is the reciprocal of Young's modulus. E can be calculated for any given direction in terms of the elastic moduli $s_{11} \ldots s_{66}$ (which are constants representing (strain)/(stress)) and the angles between the given direction and the crystal axes. If $\cos^{-1}\alpha$, $\cos^{-1}\beta$, $\cos^{-1}\gamma$ be the angles between this direction and the O$\breve{a}$, O$\bar{b}$, Oc axes respectively, then it can be shown that, for crystals belonging to the rhombic system, E is given [14] by

$$E = s_{11}\alpha^4 + s_{22}\beta^4 + s_{33}\gamma^4 + (s_{44}+2s_{23})\beta^2\gamma^2 + (s_{55}+2s_{31})\gamma^2\alpha^2 + (s_{66}+2s_{12})\alpha^2\beta^2. \quad . \quad . \quad (4.1a)$$

If the extension moduli for the directions of the length of resonators Type I., II., and III. be denoted by E_1, E_2, and E_3 respectively, then we have

$$\left.\begin{aligned} E_1 &= \tfrac{1}{4}(s_{11}+s_{22}+s_{66}+2s_{12}), \\ E_2 &= \tfrac{1}{4}(s_{22}+s_{33}+s_{44}+2s_{23}), \\ E_3 &= \tfrac{1}{4}(s_{11}+s_{33}+s_{55}+2s_{13}). \end{aligned}\right\} \quad . \quad . \quad (4.1b)$$

It is seen that in equation (4.1) all quantities are known except n_0 and E; hence a knowlege of E, which can be obtained from statical experiments [2], leads to the value

of n_0, or, conversely, a determination of n_0 may be regarded as a means of determining E.

The simple expression (4.1) requires, however, three correction factors. These are corrections due to

(i.) The finite lateral dimensions of the crystal bar.

(ii.) The secondary strains due to the charges produced on the electrodes by the vibrating bar on account of its piezo-electric properties.

(iii.) The adiabatic conditions which hold in a dynamical experiment as opposed to the sensibly isothermal conditions holding in a statical experiment.

(i.) *Corrections due to finite lateral dimensions of the crystal bar.*

The formula (4.1) is based on the fundamental assumption that the motion of any particle in the bar is purely longitudinal, and that the effect of lateral motion is negligible. This assumption can be considered to hold when the lateral dimensions of the bar are negligible in comparison with the longitudinal dimensions and when k, the order of the mode of vibration, is small.

Rayleigh [15] has shown how to allow for the effect of lateral motion by obtaining a second approximation for the frequency in the case of an isotropic bar of circular cross-section.

If T_0 be the kinetic energy of the bar vibrating without lateral motion, $T_0+\delta T_0$ the kinetic energy when lateral motion is permitted, and n_0 and n the corresponding frequencies, then

$$\frac{n}{n_0}=\sqrt{\frac{T_0}{T_0+\delta T_0}}\doteq 1-\frac{1}{2}\frac{\delta T_0}{T_0} \quad . \quad . \quad . \quad (4.2)$$

when $\dfrac{\delta T_0}{T_0}$ is small.

For the case of a circular rod of radius r, length l, for which Poisson's ratio is μ he has shown that

$$\frac{n}{n_0}=1-\frac{k^2\mu^2\pi^2}{4}\cdot\frac{r^2}{l^2}. \quad . \quad . \quad . \quad . \quad (4.3)$$

It will be noted that the correction term depends on k and r/l, and may, therefore, be important in the case of piezo-electric resonators in which the lateral dimensions are usually comparable with the length.

The correction for quartz resonators has been stated, without proof, by v. Laue[16] and Giebe and Scheibe[17]. We shall now calculate this correction for the general case of a rectangular parallelepiped of crystalline material.

With the centre of the parallelepiped as origin, take a set of rectangular axes Ox, Oy, Oz parallel to the edges. Let the length of the bar be L_1 and let it be parallel to the Ox axis; let the lengths of the edges parallel to Oy and Oz be $2a_1$ and $2b_1$ respectively.

If u, v, w be the displacements parallel to the coordinate axes, then the kinetic energy, T, of the vibrating bar is given, in the usual notation, by

$$T = \frac{1}{2}\rho \iiint (\dot{u}^2 + \dot{v}^2 + \dot{w}^2)\, dx\, dy\, dz, \quad . \quad . \quad (4.4)$$

the integration being taken over the whole volume of the bar, *i. e.*, between the limits $x = \pm \frac{L_1}{2}$, $y = \pm a_1$, $z = \pm b_1$.

For the case of vibration parallel to Ox, for an infinitely thin bar, this equation reduces to

$$T_0 = \frac{1}{2}\rho \iiint \dot{u}^2\, dx\, dy\, dz. \quad . \quad . \quad . \quad (4.5)$$

Now, for longitudinal vibrations, the displacement u may be written in the form

$$\left.\begin{aligned} u &= A \cos\frac{k\pi x}{L_1}, \\ \text{whence} \qquad \dot{u} &= \dot{A} \cos\frac{k\pi x}{L_1}, \end{aligned}\right\} \quad . \quad . \quad . \quad . \quad . \quad (4.6)$$

where A is a function of time only.

Substitution of (4.6) in (4.5) and integration gives

$$T_0 = \tfrac{1}{4}\rho \dot{A}^2 \omega L_1, \quad . \quad . \quad . \quad . \quad . \quad (4.7)$$

where ω is the cross-sectional area of the bar.

We have next to calculate the additional kinetic energy, δT_0, which arises when the bar is not infinitely thin. If the bar be subject to a static tension X_x in the direction of Ox, then it can be assumed that, in general, the displacements are given by equations of the type

$$\left.\begin{aligned} u &= (Px + Qy + Rz)X_x, \\ v &= Myx_x, \\ w &= Nzx_x, \end{aligned}\right\} \quad . \quad . \quad . \quad (4.8)$$

where P, Q, R, M, and N are constants,—elastic moduli in Voigt's nomenclature.

For an infinitely thin bar these reduce to the single equation $u = \mathrm{P}x\mathrm{X}_x$, whence

$$\mathrm{X}_x = \frac{1}{\mathrm{P}}\frac{\partial u}{\partial x}. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (4.9)$$

In the dynamical problem of the vibration of a thick bar, we may assume that the displacements are of the form given by equation (4.8), and that, to a first approximation, X_x is given by equation (4.9). We can, therefore, write the following equations for the *extra* velocities $\delta\dot{u}$, $\delta\dot{v}$, $\delta\dot{w}$, arising when the bar is not infinitely thin.

$$\left.\begin{aligned} \delta\dot{u} &= \frac{1}{\mathrm{P}}(\mathrm{Q}y + \mathrm{R}z)\frac{\partial^2 u}{\partial x\,\partial t}, \\ \delta\dot{v} &= \frac{\mathrm{M}y}{\mathrm{P}}\frac{\partial^2 u}{\partial x\,\partial t}, \\ \delta\dot{w} &= \frac{\mathrm{N}z}{\mathrm{P}}\frac{\partial^2 u}{\partial x\,\partial t}. \end{aligned}\right\} \quad . \quad . \quad . \quad (4.10)$$

The kinetic energy of a volume element $dx\,dy\,dz$ due to vibration parallel to Ox is given by

$$\tfrac{1}{2}\rho(\dot{u}+\delta\dot{u})^2 dx\,dy\,dz = \tfrac{1}{2}\rho\dot{u}^2 dx\,dy\,dz + \rho\dot{u}\,\delta\dot{u}\,dx\,dy\,dz + \tfrac{1}{2}\rho(\delta\dot{u})^2 dx\,dy\,dz.$$

On integration over the volume of the bar, the first term of this expression gives T_0 of equation (4.7)—the kinetic energy in the absence of lateral motion; the integral of the second term vanishes, since the integral of the factor $(\mathrm{Q}y + \mathrm{R}z)$ taken over the cross-section of the bar is zero. Assuming that the displacement u is given, to a first approximation, by equation (4.6), the increase, $\delta\mathrm{T}_x$, in the kinetic energy, due to the extra vibration parallel to Ox, is thus given by

$$\begin{aligned} \delta\mathrm{T}_x &= \frac{1}{2}\rho\iiint(\delta\dot{u})^2 dx\,dy\,dz \\ &= \frac{1}{2}\rho\dot{\mathrm{A}}^2\frac{k^2\pi^2}{\mathrm{P}^2\mathrm{L}_1^2}\iiint(\mathrm{Q}y+\mathrm{R}z)^2\sin^2\frac{k\pi x}{\mathrm{L}_1}\,dx\,dy\,dz, \end{aligned}$$

whence

$$\delta\mathrm{T}_x = \frac{1}{12}\rho\dot{\mathrm{A}}^2\cdot\frac{\omega}{\mathrm{L}_1}\cdot\frac{k^2\pi^2}{\mathrm{P}^2}(\mathrm{Q}^2a_1^2 + \mathrm{R}^2b_1^2). \quad . \quad (4.11)$$

Similar expressions hold for δT_y and δT_z, the kinetic energies due to the extra vibration parallel to Oy and Oz. Finally, we obtain for δT_0, the kinetic energy due to extra vibration in all three directions, the expression

$$\delta T_0 = \frac{1}{12}\rho \dot{A}^2 \frac{\omega}{L_1} k^2\pi^2 \left\{ \frac{(Q^2+M^2)a_1^2+(R^2+N^2)b_1^2}{P^2} \right\}, \quad . \quad . \quad . \quad (4.12)$$

whence, by equation (4.2),

$$\frac{n}{n_0} = 1 - \frac{k^2\pi^2}{6} \left\{ \frac{(Q^2+M^2)a_1^2+(R^2+N^2)b_1^2}{P^2L_1^2} \right\}. \quad (4.13)$$

It remains to evaluate the constants P, Q, R, M, and N in this formula for the three types of resonators employed.

Consider a crystalline medium subject to a system of stresses, and suppose that the coordinate axes coincide with the crystal axes ; if the components of stress be denoted by $X_x \ldots X_y$, then the resulting components of strain $x_x \ldots x_y$ are, in general, given [18] by six equations of the type

$$\left.\begin{aligned} -x_x &= s_{11}X_x + s_{12}Y_y + s_{13}Z_z + s_{14}Y_z + s_{15}Z_x + s_{16}X_y, \\ &\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ -x_y &= s_{61}X_x + s_{62}Y_y + s_{63}Z_z + s_{64}Y_z + s_{65}Z_x + s_{66}X_y, \end{aligned}\right\} \quad . \quad . \quad . \quad (4.14)$$

where $s_{11} \ldots s_{66}$ are the elastic moduli. For crystals of the rhombic hemihedral class only nine of the moduli differ from zero.

If the coordinate axes to which the components of stress and strain are referred do not coincide with the crystal axes, but are connected together by the transformation given in (3.3), then the components of stress and strain ($X'_x \ldots$, $x'_x \ldots$) satisfy six similar equations of the type

$$-x'_x = s'_{11}X'_x + s'_{12}Y'_y + s'_{13}Z'_z + s'_{14}Y'_z + s'_{15}Z'_x + s'_{16}X'_y, \quad (4.15)$$

where the effective elastic moduli $s'_{11} \ldots s'_{66}$ are calculable in terms of the moduli $s_{11} \ldots s_{66}$ and the direction-cosines $\alpha_1 \ldots \gamma_3$ [19].

For a Type I. resonator, it has been shown that the components of stress reduce to two—X'_x and Y'_y—in the direction of the Ox' and Oy' axes, when the applied field is parallel to the Oz or Oz' axis. In the case of a static electric field, the two components of stress are equal in magnitude; in the case of a piezo-electric resonator, however,

the stresses only become appreciable in the neighbourhood of resonance, and since resonance for vibrations parallel to Ox' and Oy' will scarcely ever occur simultaneously, we shall be justified in neglecting Y'_y in comparison with X'_x for vibrations parallel to Ox', and X'_x in comparison with Y'_y for vibrations parallel to Oy'. This simplifies the equations (4.15) considerably; further simplification arises from the fact that the effective elastic moduli s'_{41}, s'_{42}, s'_{51}, and s'_{52} vanish for a resonator of this type. Introducing these modifications, equations (4.15) reduce to

$$\left.\begin{aligned} -x'_x = s'_{11}X'_x; \quad -y'_y = s'_{21}X'_x; \quad -z'_z = s'_{31}X'_x; \\ -x'_y = s'_{61}X'_x; \quad z'_x = y'_z = 0; \end{aligned}\right\} \quad (4.16)$$

for vibrations parallel to Ox', and similar equations hold for vibrations parallel to Oy'.

The effective elastic moduli are easily calculated, remembering that

$$\alpha_1 = \alpha_2 = \beta_2 = \frac{1}{\sqrt{2}}; \quad \beta_1 = -\frac{1}{\sqrt{2}}; \quad \gamma_3 = 1;$$
$$\alpha_3 = \beta_3 = \gamma_1 = \gamma_2 = 0.$$

In the present case they have the values

$$\left.\begin{aligned} s'_{11} &= \tfrac{1}{4}(s_{11} + s_{22} + s_{66} + 2s_{12}), \\ s'_{21} &= \tfrac{1}{4}(s_{11} + s_{22} - s_{66} + 2s_{12}), \\ s'_{31} &= \tfrac{1}{2}(s_{13} + s_{23}), \\ s'_{61} &= -\tfrac{1}{2}(s_{11} - s_{22}). \end{aligned}\right\} \quad . \quad . \quad . \quad (4.17)$$

To deduce the displacements u', v', w' (parallel to Ox', Oy', and Oz' respectively), from the components of strain, we use the well-known equations

$$\left.\begin{aligned} x'_x = \frac{\partial u'}{\partial x'}; \quad y'_y = \frac{\partial v'}{\partial y'}; \quad z'_z = \frac{\partial z'}{\partial w'}; \\ y'_z = z'_y = \frac{\partial v'}{\partial z'} + \frac{\partial w'}{\partial y'}; \quad z'_x = x'_z = \frac{\partial w'}{\partial x'} + \frac{\partial u'}{\partial z'}; \\ x'_y = y'_x = \frac{\partial u'}{\partial y'} + \frac{\partial v'}{\partial x'}; \end{aligned}\right\} \quad (4.18)$$

together with the conditions that, when the vibration is parallel to Ox', there is no rotation about this axis, neither is the axis displaced.

These conditions lead to the equations

$$\frac{\partial w'}{\partial y'} - \frac{\partial v'}{\partial z'} = 0; \quad \frac{\partial v'}{\partial x'} = \frac{\partial w'}{\partial x'} = 0. \quad . \quad . \quad (4.19)$$

These equations, together with (4.16), lead to

$$\left.\begin{aligned} -\frac{\partial u'}{\partial x'} = s'_{11}X'_x; \quad -\frac{\partial v'}{\partial y'} = s'_{21}X'_x; \quad -\frac{\partial w'}{\partial z'} = s'_{31}X'_x; \\ \frac{\partial u'}{\partial y'} = s'_{61}X'_x; \\ \frac{\partial u'}{\partial z'} = \frac{\partial v'}{\partial x'} = \frac{\partial v'}{\partial z'} = \frac{\partial w'}{\partial x'} = \frac{\partial w'}{\partial y'} = 0; \end{aligned}\right\} \quad . \quad . \quad . \quad (4.20)$$

whence

$$-u' = (s'_{11}x' + s'_{61}y')X'_x; \quad -v' = s'_{21}y'X'_x; \quad -w' = s'_{31}z'X'_x. \quad . \quad . \quad . \quad (4.21)$$

Comparing equations (4.21) and (4.8), and remembering that the present Ox', Oy', Oz' axes are identical with the Ox, Oy, Oz axes of (4·8), we see that

$$P = -s'_{11}; \quad Q = -s'_{61}; \quad R = 0; \quad M = -s'_{21}; \quad N = -s'_{31}.$$

Hence, from equation (4.13),

$$\frac{n}{n_0} = 1 - \frac{k^2\pi^2}{6}\left\{\frac{(s'^2_{61} + s'^2_{21})a_1^2 + s'^2_{31}b_1^2}{s'^2_{11}L_1^2}\right\}. \quad . \quad (4.22)$$

Similar expressions hold for resonators of Type II. and Type III.

For a Type II. resonator, whose length, L_2, is at 45° to the $O\bar{b}$ axis, whose width, $2a_2$, is at 45° to the Oc axis, and whose depth, $2b_z$, is parallel to the $Oă$ axis, it may be shown that

$$\left.\begin{aligned} \frac{n}{n_0} = 1 - \frac{k^2\pi^2}{6}\left\{\frac{(s''^2_{32} + s''^2_{42})a_2^2 + s''^2_{12}b_2^2}{s''^2_{22}L_2^2}\right\}, \\ \text{where} \\ s''_{22} = \tfrac{1}{4}(s_{22} + s_{33} + s_{44} + 2s_{23}), \\ s''_{32} = \tfrac{1}{4}(s_{22} + s_{33} - s_{44} + 2s_{23}), \\ s''_{42} = \tfrac{1}{2}(s_{33} - s_{22}), \\ s''_{12} = \tfrac{1}{2}(s_{12} + s_{13}). \end{aligned}\right\} \quad . \quad (4.23)$$

For a Type III. resonator, whose length, L_3, is at 45° to the Oc axis, whose width, $2a_3$, is at 45° to the $O\breve{a}$ axis, and whose depth, $2b_3$, is parallel to the $O\breve{b}$ axis, it may be shown that

$$\left.\begin{aligned}\frac{n}{n_0}&=1-\frac{k^2\pi^2}{6}\left\{\frac{(s'''_{13}+s'''_{35})a_3^2+s'''_{23}b_3^2}{s''''_{33}L_3^2}\right\},\\ \text{where}\quad s'''_{33}&=\tfrac{1}{4}(s_{11}+s_{33}+s_{55}+2s_{13}),\\ s'''_{13}&=\tfrac{1}{4}(s_{11}+s_{33}-s_{55}+2s_{13}),\\ s'''_{35}&=\tfrac{1}{2}(s_{11}-s_{33}),\\ s'''_{23}&=\tfrac{1}{2}(s_{12}+s_{23}).\end{aligned}\right\}\quad(4.24)$$

Inserting the values of the elastic moduli determined by Mandell[2], we find that

$$\left.\begin{aligned}\frac{n}{n_0}&=1-k^2\left[0{\cdot}09449\frac{a_1^2}{L_1^2}+0{\cdot}007998\frac{b_1^2}{L_1^2}\right]\\ &\equiv 1-k^2\phi_1\ (\text{say})\ \text{for Type I. resonators},\\ \frac{n}{n_0}&=1-k^2\left[0{\cdot}07865\frac{a_2^2}{L_2^2}+0{\cdot}2436\frac{b_2^2}{L_2^2}\right]\\ &\equiv 1-k^2\phi_2\ (\text{say})\ \text{for Type II. resonators},\\ \frac{n}{n_0}&=1-k^2\left[1{\cdot}11\frac{a_3^2}{L_3^2}+0{\cdot}004539\frac{b_3^2}{L_3^2}\right]\\ &\equiv 1-k^2\phi_3\ (\text{say})\ \text{for Type III. resonators}.\end{aligned}\right\}\quad\ldots\quad(4.25)$$

The following table gives details of the resonators used in the experiments, together with the value of the appropriate ϕ calculated from (4.25), and $\frac{n}{n_0}$ for the fundamental mode of vibration ($k=1$).

TABLE I.

Resonator no.	Type.	Length, L (cm.).	Breadth, $2a$ (cm.).	Thickness, $2b$ (cm.).	ϕ.	$\frac{n}{n_0}$ for $k=1$.
I.a	I.	1·892	0·630	0·164	$\phi_1=0{\cdot}00263_6$	$0\,997_4$
I.b	I.	1·343	0·497	0·174	$\phi_1=0{\cdot}00325_1$	$0{\cdot}996_7$
II.a.....	II.	1·632	0·539	0·149	$\phi_2=0{\cdot}00266_1$	$0{\cdot}997_3$
III.a	III.	1·333	0·381	0·194	$\phi_3=0{\cdot}0228_7$	$0{\cdot}997_1$
III.b	III.	1·290	0·624	0·150	$\phi_3=0{\cdot}0650_2$	$0{\cdot}935_0$

It is seen that, whilst this correction is comparatively small for Type I. and Type II. resonators, it becomes appreciable in the case of Type III. resonators; the correction increases also with the order of vibration, as is evident from the fact that in the case of resonators mounted in the way described $k=3$ for the first overtone, $k=5$ for the second overtone, and so on [16].

(ii.) *Correction due to secondary charges and strains.*

A static electric field, applied perpendicular to the two main faces of a piezo-electric resonator, will, in virtue of the converse piezo-electric effect, produce linear strains parallel to the edges bounding these faces. These strains, in virtue of the direct piezo-electric effect, will give rise to secondary charges on the electrodes, and hence to a secondary field opposing the applied field. Finally, the secondary field produces secondary strains opposite in direction to the original strains. It is clear that this secondary effect is equivalent to a decrease in the extension modulus, and the appropriate correction has been calculated by Giebe and Scheibe [17].

This argument, however, applies only under statical conditions. For the dynamical conditions holding in a piezo-electric resonator, Giebe and Scheibe [17] have shown, both experimentally and theoretically, that two different cases are possible. In the one case the charges produced during vibration are *surface* charges, whilst in the other they are *volume* charges. In the former case—surface charges—the secondary field is appreciable, and the correction approximates to that for the statical case. In the latter case—volume charges—the secondary field, and, therefore, the correction, is negligible. An argument similar to that employed by Vigoureux [20] shows that the charges produced in the resonators used in these experiments are *volume* charges, and hence the correction can be neglected.

(iii.) *Correction for adiabatic conditions.*

Voigt [21] has shown how to derive the relationship between an isothermal extension modulus E and the corresponding adiabatic extension modulus E_s at absolute temperature T.

If α be the coefficient of linear expansion in the direction in which E is measured, ρ be the density of the crystal, C its specific heat, and J the mechanical equivalent of heat, then

$$E_s - E = -\frac{\alpha^2 T}{\rho J C}. \quad . \quad . \quad . \quad . \quad (4.26)$$

The values of ρ, C, α, $E_s'-E$, E, and E_s for the various types of resonators at a temperature of 15° C. are tabulated below, E being calculated from Mandell's results [2].

TABLE II.

$\rho=1{\cdot}766$; $C=0{\cdot}328$.

Resonator type.	α in direction of length.	E_s-E (cm.²/dyne).	E (cm.²/dyne).	E_s (cm.²/dyne).
I.	$4{\cdot}9_0\times10^{-5}$	$-2{\cdot}8_5\times10^{-14}$	$3{\cdot}58\times10^{-12}$	$3{\cdot}55_1\times10^{-12}$
II.	$4{\cdot}1_5\times10^{-5}$	$-2{\cdot}0_4\times10^{-14}$	$3{\cdot}86_5\times10^{-12}$	$3{\cdot}84_5\times10^{-12}$
III.	$5{\cdot}2_4\times10^{-5}$	$-3{\cdot}2_5\times10^{-14}$	$8{\cdot}44\times10^{-12}$	$8{\cdot}40_7\times10^{-12}$

It is seen that this correction is less than 1 per cent. for all three types of resonators.

Below are tabulated the calculated natural frequencies of the five resonators for their fundamental mode of vibration. The fifth column gives the frequency, n_0, calculated from (4.1), using the adiabatic values, E_s, of the extension moduli; the sixth the frequency, n, given in the sixth column, allows for the finite lateral dimensions of the bars, using the factors given in Table I.; the last column gives λ, the electromagnetic wave-length (in metres) corresponding to n.

TABLE III.

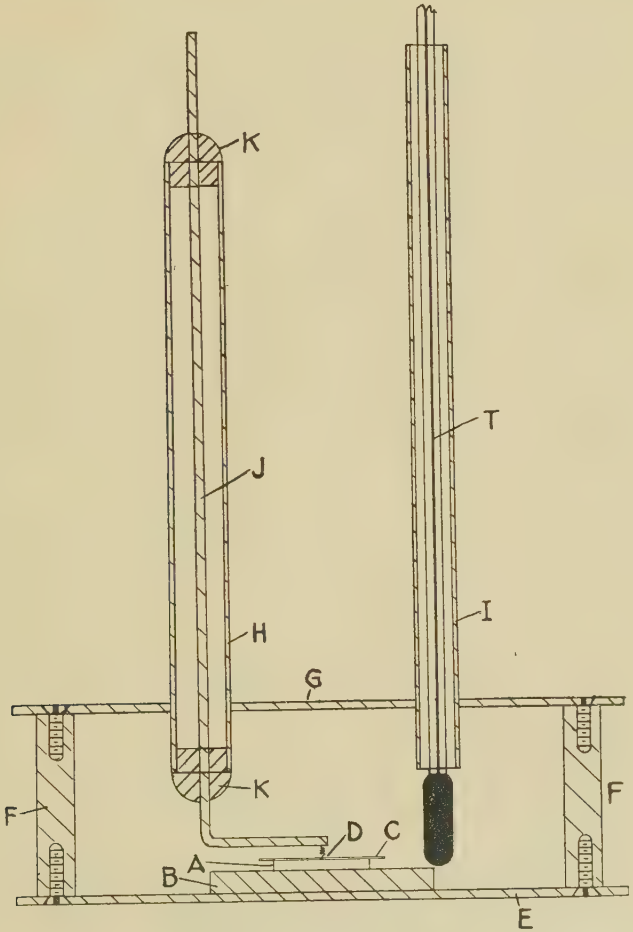
Resonator no.	Type.	L (cm.).	E_s(cm.²/dyne).	n_0.	n.	λ (metres).
I. *a*.....	I.	1·892	$3{\cdot}55_1\times10^{-12}$	$1{\cdot}055\times10^5$	$1{\cdot}053\times10^5$	2848
I. *b*.....	I.	1·343	$3{\cdot}55_1\times10^{-12}$	$1{\cdot}487\times10^5$	$1{\cdot}482\times10^5$	2023
II. *a*	II.	1·632	$3{\cdot}84_5\times10^{-12}$	$1{\cdot}175\times10^5$	$1{\cdot}172\times10^5$	2557
III. *a* ...	III.	1·333	$8{\cdot}40_7\times10^{-12}$	$0{\cdot}9736\times10^5$	$0{\cdot}9513\times10^5$	3151
III. *b*.....	III.	1·290	$8{\cdot}40_7\times10^{-12}$	$1{\cdot}006\times10^5$	$0{\cdot}9406\times10^5$	3187

§ 5. *Experimental Details.*

In an experiment which involves the measurement of the response frequency of a resonator at different temperatures, the method of mounting the resonator is of importance, because of the variation of response frequency with the magnitude of the air-gap. Unless proper precautions are taken, the effect due to the change in the magnitude of the air-gap with temperature may mask the effect due to a change in the extension modulus with temperature. Various devices were tried, and the one finally adopted is shown in

fig. 2. The crystal bar A rests on a plane circular brass disk, B, about 3 cm. in diameter ; B forms one electrode of the resonator. The other electrode consists of a light, thin

Fig. 2.



Resonator holder.

sheet of mica, C, coated on both sides with tinfoil, to which is soldered one end of a short length, D, of fine copper wire of 0·0035 cm. diameter. Contact can be made between this electrode and the external circuit by soldering the free end of D to a suitable conductor ; it was found that the resonator

would vibrate freely if sufficient "slack" were left in D. Using a thicker wire, or neglecting to allow sufficient "slack" in D, prevents free vibration of the resonator. In all cases the area of the mica-tinfoil sheet, C, was greater than the area of the bar A, in order to eliminate any possible edge-effects.

The disk B was soldered to a larger brass disk, E, 9 cm. in diameter; E carried three ebonite rods, F, which supported a brass disk, G, at a distance of 6 cm. above E. The disk G carried two brass tubes, H and I, the latter serving as a guide tube for the thermometer, T; I is placed so as to bring the thermometer-bulb as near as possible to A.

A brass rod, J, passed down the centre of the tube H, and two ebonite plugs, K, served to insulate J from H. The rod J is bent at right angles near its lower end, and the free end of the thin wire D soldered to it. It will be seen from fig. 1 that one electrode of the piezo-resonator is earthed, whilst the other is insulated and connected to a mercury-cup key, which, in turn, is connected to the insulated plates of the wave-meter condenser. The brass disk B is the earthed electrode in the apparatus shown in fig. 2, whilst C is the insulated electrode, the upper end of J being connected to the mercury-cup key.

To obtain reliable results it is essential to arrange that the temperature of the resonator has been constant for an hour or two before taking a reading. For this purpose the above apparatus is placed in a double-walled tin vessel, open at one end and well lagged with heat-insulating material; the space between the apparatus and the inner wall of the vessel is packed with cotton-wool to prevent convection currents, and the lid of the vessel is fixed on with wax to obtain an air-tight container.

During preliminary experiments trouble was caused by condensation of water on the resonator after a change of temperature; this resulted in the crystal bar becoming fixed on the lower electrode and refusing to function as a resonator. To avoid this, a vessel containing phosphorus pentoxide was placed near the resonator, and the cotton-wool was heated, before use, to drive off any moisture it might contain.

The resonator could be brought to any desired temperature by filling the space between the walls of the tin vessel with water, ice, or freezing mixtures; above room-temperatures radiation losses were compensated by a small electrically-heated coil beneath the vessel.

The resonator and its holder were screened electrostatically by earthing the tin vessel and the lid; the mercury key

and leads were enclosed in an earthed cage for the same purpose.

The principle of the method used for determining the resonant frequency, n, of the resonators has been described above. In practice the wave-meter condenser is set so as to obtain a resonance curve showing the characteristic crevasse when the resonator is in circuit. The tuning condenser of the valve-oscillator is then adjusted until the deflexion of the galvanometer G of fig. 1 is a minimum—corresponding to the lowest point in the crevasse. A measurement of the frequency or wave-length of the valve-oscillator then gives the resonant frequency of the resonator. For this purpose the resonator is disconnected by means of the key K of fig. 1, and the wave-meter condenser adjusted so as to give maximum galvanometer deflexion; the wave-length of the oscillator can then be read directly on the wave-meter scale, if desired. In practice it was found more desirable to utilize the angular scale and vernier with which the wave-meter condenser is provided. A large-scale graph was drawn connecting condenser readings and wave-lengths, the points on this graph being determined by finding the condenser readings corresponding to the wave-lengths given in the N.P.L. calibration chart of the instrument; the wave-length corresponding to any particular condenser reading could then be determined from this graph. By this method fairly accurate absolute values of wave-lengths can be obtained, whilst the change in wave-length corresponding to a given change in condenser reading is given much more accurately.

The accuracy of the above method of determining resonant frequencies is determined almost entirely by the accuracy of the wave-meter, which is of the order of 1 in 1000. It is probable that the accuracy of the results for change of wave-length (or frequency) with temperature is greater than this figure when using the above method of interpolation, since the absolute accuracy of the wave-meter condenser is inferior to its conformity to law.

It should be stated that the minimum deflexion corresponding to the bottom of the crevasse was sufficiently well defined in most cases to enable the corresponding wave-length to be read to within 0·05° on the vernier scale. At 4000 m. 0·05° on the condenser scale corresponds to 0·9 m.; at 3000 m. to 0·5 m.; at 2000 m. to 0·4 m.; at 1000 m. to 0·2 m.; at 700 m. to 0·12 m. Thus a change of 1 in 5000 in the resonant frequency could be detected in most cases.

With the crevasse method of determining resonant frequencies it is first necessary to find the frequencies

approximately, in order to know where to set the wave-meter condenser. This can be done by connecting the resonator in parallel with the tuning condenser of a valve oscillator of the regenerative type, the oscillatory circuit being between grid and filament. A pair of phones is connected in the anode circuit, and, on swinging the tuning condenser, a characteristic click is heard in the phones whenever the frequency of the oscillatory circuit becomes equal to a resonant frequency of the piezo-electric resonator.

§6. *Experimental Results at Room Temperature.*

The results obtained at a temperature of 15° C. are given in the following table. The third and fourth columns give the resonant frequency and wave-length, n and λ (calculated above), for a temperature of 15° C., whilst the last columns give the resonant frequency and wave-length, n_e and λ_e, as determined by experiment at this temperature.

Table IV.

Resonator no.	Type.	n.	λ (metres).	n_e.	λ_e (metres).
I. *a*	I.	$1{\cdot}053\times10^5$	2848	$0{\cdot}9940\times10^5$	$3015{\cdot}_9$
I. *b*	I.	$1{\cdot}482\times10^5$	2023	$1{\cdot}404\times10^5$	$2134{\cdot}_5$
II. *a*	II.	$1{\cdot}172\times10^5$	2557	$1{\cdot}268\times10^5$	$2364{\cdot}_9$
III. *a*.....	III.	$0{\cdot}9513\times10^5$	3151	$0{\cdot}9114\times10^5$	$3289{\cdot}_1$
III. *b*	III.	$0{\cdot}9406\times10^5$	3187	$0{\cdot}9101\times10^5$	$3294{\cdot}_0$

All the values given in this table relate to the fundamental mode of vibration of the bars ($k=1$). It is seen that the calculated and experimental values show considerable discrepancies; to elucidate these discrepancies it is of interest to reverse the calculation, and to use the experimental results to derive values for the extension moduli.

To do this, the value, n_e, of the resonant frequency obtained by experiment is corrected in each instance for the finite lateral dimensions of the bar by division by the factors given in Table I.; the value of the frequency so obtained corresponds to an infinitely thin bar, and is denoted by n_{0e}. The adiabatic extension moduli are then obtained by substitution in (4.1), these values being denoted by E_d. The values of λ_e, n_e, n_{0e}, and E_d obtained in this way are tabulated below; the corresponding adiabatic moduli, E_s, calculated from Mandell's statical experiments are also given for purposes of comparison.

TABLE V.

Resonator no.	Type.	λ_e (metres).	n_e.	n_{0e}.	E_d (cm.²/dyne).	E_s (cm.²/dyne).
I. *a*.....	I.	$3015{\cdot}_9$	$0{\cdot}9940\times10^5$	$0{\cdot}9966\times10^5$	$3{\cdot}98_2\times10^{-12}$	$3{\cdot}55_1\times10^{-12}$
I. *b*.....	I.	$2134{\cdot}_5$	$1{\cdot}404\times10^5$	$1{\cdot}409\times10^5$	$3{\cdot}95_3\times10^{-12}$	$3{\cdot}55_1\times10^{-12}$
II. *a*	II.	$2364{\cdot}_9$	$1{\cdot}268\times10^5$	$1{\cdot}271\times10^5$	$3{\cdot}28_9\times10^{-12}$	$3{\cdot}84_5\times10^{-12}$
III. *a*.....	III.	$3289{\cdot}_1$	$0{\cdot}9114\times10^5$	$0{\cdot}9328\times10^5$	$10{\cdot}9_2\times10^{-12}$	$8{\cdot}40_7\times10^{-12}$
III. *b*.....	III.	$3294{\cdot}_0$	$0{\cdot}9101\times10^5$	$0{\cdot}9734\times10^5$	$11{\cdot}1_4\times10^{-12}$	$8{\cdot}40_7\times10^{-12}$

It is seen that the values of E_d derived from resonators I. *a*, I. *b* and III. *a*, III. *b* are in satisfactory agreement, showing that this method does lead to consistent results. The differences between the experimental values are of the order of 1 per cent. and 2 per cent. respectively. The discrepancies between E_d and E_s are much greater, and it is, therefore, of interest to examine the possible errors involved in the determination of E by the dynamical method described in this paper.

The first source of error lies in the measurement of the resonant wave-length, λ_e; the error due to this cause has been discussed in §5, where it is shown that λ_e can be determined to within 1 in 1000; this would lead to an uncertainty of 1 in 500 in E, since E depends on the square of λ_e.

A second possible source of error is non-uniformity in the lengths of the resonators and errors in the actual measurement of these lengths. Errors due to this cause can be estimated at 0·5 per cent. at the outside; this introduces a possible inaccuracy of 1 per cent. in E, since E depends on the square of the length.

A third possible source of error is that due to incorrect orientation of the length of the resonators with respect to the crystal axes; the magnitude of this error can be estimated from equation (4·1 *a*) by insertion of the appropriate values of α, β, γ. Assuming the values given by Mandell[2], it can be calculated that an error of 1° in the orientation introduces an error in E of about 1 per cent. for Type I. resonators, and about 0·5 per cent. for Type II. and Type III. resonators. If we assume a possible uncertainty of 2° on either side of the correct direction, this would lead to an uncertainty of 4 per cent. in E for Type I. resonators and 2 per cent. for Type II. and Type III. resonators.

Comparison with the experimental results given in Table V. shows that the differences in the two values of

E_s obtained for resonators I. *a*, I. *b*, and III. *a*, III. *b* are satisfactorily acounted for, and that these differences are of the order of magnitude indicated by the discussion of the possible sources of error.

The following formulæ may be used to pre-determine the approximate resonant frequency, n_0, of the three types of resonators at a temperature of 15° C. ; they are based on the experimental values obtained for E_d, and are valid only for resonators whose lateral dimensions are small in comparison with their length, L.

$$\left.\begin{aligned} &\text{Resonator Type I.;} && n_0 = \frac{1{\cdot}89\times10^5}{L},\\ &\text{,, Type II.;} && n_0 = \frac{2{\cdot}07\times10^5}{L},\\ &\text{,, Type III.;} && n_0 = \frac{1{\cdot}13\times10^5}{L}. \end{aligned}\right\} \quad . \quad (6.1)^*$$

§ 7. *Experimental Results for the Temperature Variation of Resonant Wave-length and Extension Moduli.*

The temperature variation of resonant wave-length was investigated for resonators I. *a*, II. *a*, III. *a*, representing the three different types of resonator.

The curves reproduced show the variation of the resonant wave-lengths, λ_e (in metres), with temperature, t (° C.) ; the former are taken as ordinates, the latter as abscissæ.

In terms of λ_e, L, ρ the extension modulus, E, is given by

$$E = \frac{\lambda_e^2}{36\times10^{16}\,L^2\rho} \quad . \quad . \quad . \quad . \quad . \quad . \quad (7.1)$$

The three quantities λ, L, and ρ are functions of temperature, and their variation must be allowed for when obtaining a representation of the variation of E with temperature. If L_0 be the length of the resonator at 0° C, ρ_0 the corresponding density, Δ the coefficient of volume expansion, α the coefficient of linear expansion in the direction of the length of the resonator, then

$$E = \frac{\lambda_e^2}{36\times10^{16}\,L_0^2\rho_0}[1+(\Delta-2\alpha)t]. \quad . \quad . \quad (7.2)$$

The value of $(\Delta-2\alpha)$ for the different resonators can be calculated from data given by Valasek [22], and is given by

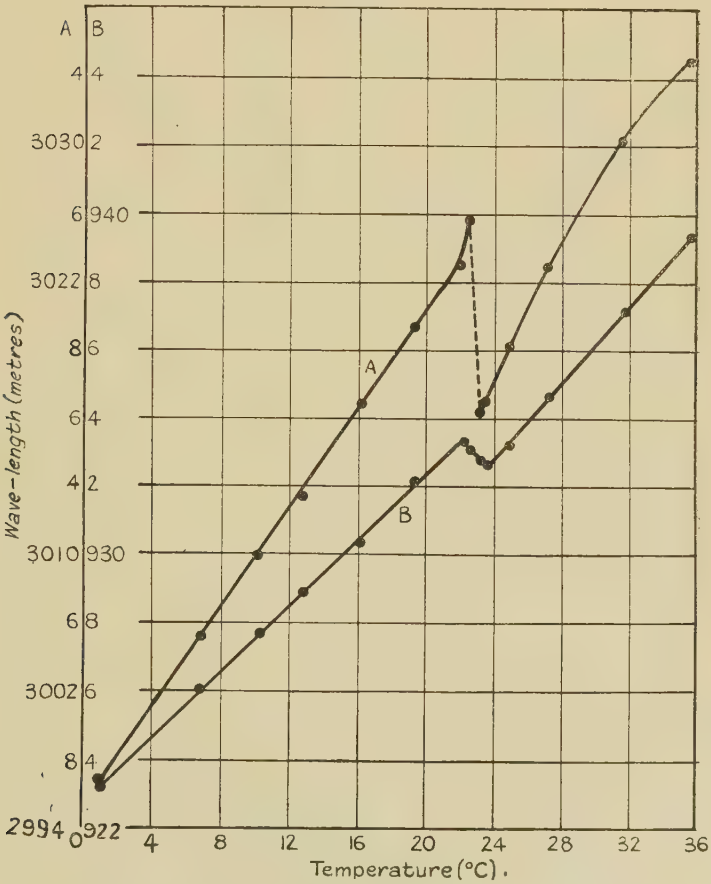
$$\begin{aligned} &\text{Resonator Type I.;} && \Delta-2\alpha = 4{\cdot}48\times10^{-5}.\\ &\text{,, ,, II.;} && \Delta-2\alpha = 6{\cdot}01\times10^{-5}.\\ &\text{,, ,, III.;} && \Delta-2\alpha = 4{\cdot}21\times10^{-5}. \end{aligned}$$

* In (6.1), L is measured in cm., n in cycles per sec.

This gives the data required to calculate the temperature variation of E when that of λ_e is given.

The temperature range for resonators I. *a*, II. *a*, and III. *a* is from 0° C. to 40° C., and for two resonance points of

Fig. 3.



Resonant wave-length–temperature curves for resonator I. *a*.

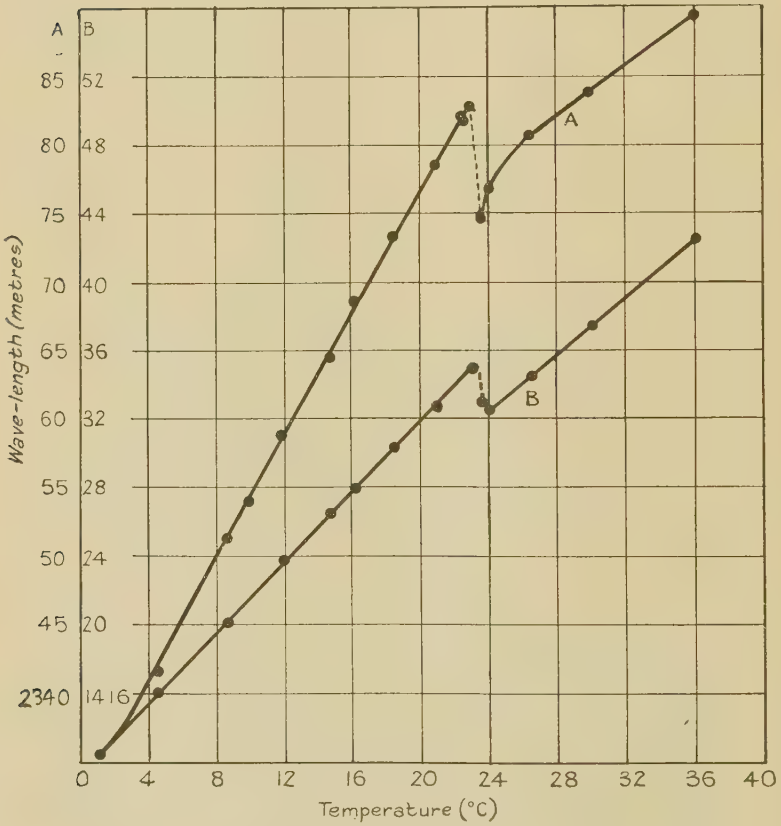
resonator II. *a* from −23° C. to 40° C. Considerable difficulty was experienced with the lower temperatures owing to the freezing of water on the ebonite plugs, thus spoiling the insulation.

The resonance points investigated were the more prominent ones, judged by their comparative effects on the resonance

curves ; for all three resonators, the number of resonance points was greater than the number studied. In some cases it was possible to identify the mode of vibration ; in others this was impossible.

Fig. 3 shows the variation of λ_e with t for resonator I. a;

Fig. 4.



Resonant wave-length-temperature curves for resonator II. a.

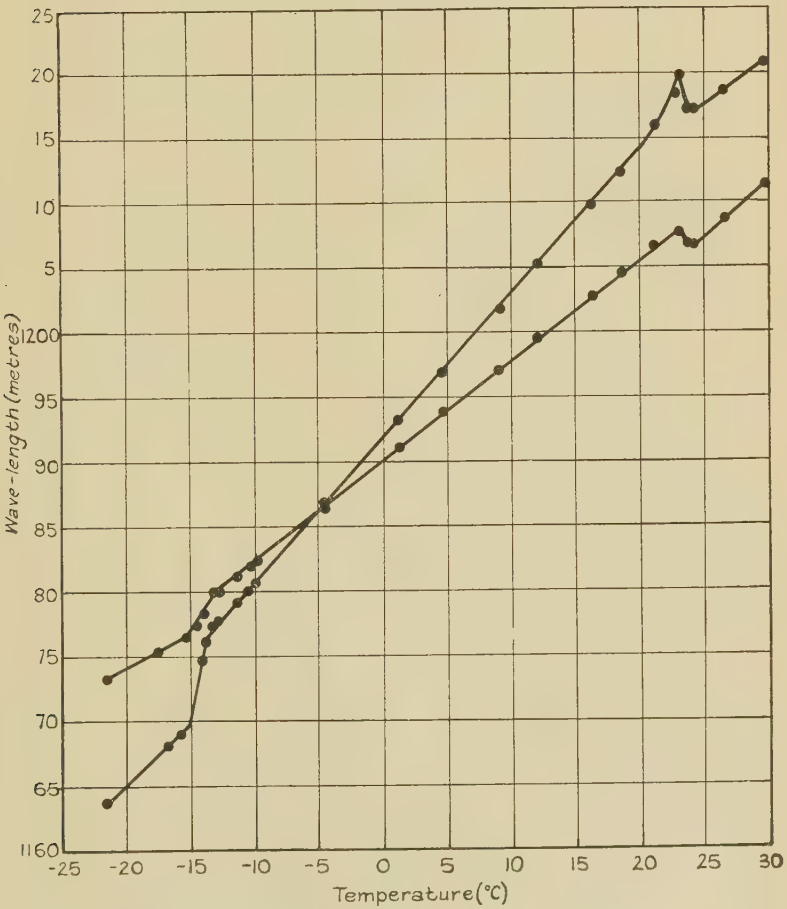
curve A relates to the fundamental mode of vibration, whilst curve B relates to the harmonic corresponding to $k=3$.

Figs. 4 and 5 show this variation for resonator II. a, whilst figs. 6 and 7 relate to resonator III. a. Curve A in figs. 4 and 6 relates to the fundamental mode of vibration, whilst the other curves relate to various harmonics.

These curves show a general similarity in the behaviour of resonators I. a and II. a ; λ_e (and consequently E)

increases regularly with temperature, the increase following a linear law. At a temperature in the neighbourhood of 23° C. there is a sudden drop in the value of λ_e and of E, accompanied by a change in the slope of the curve. Fig. 5

Fig. 5.

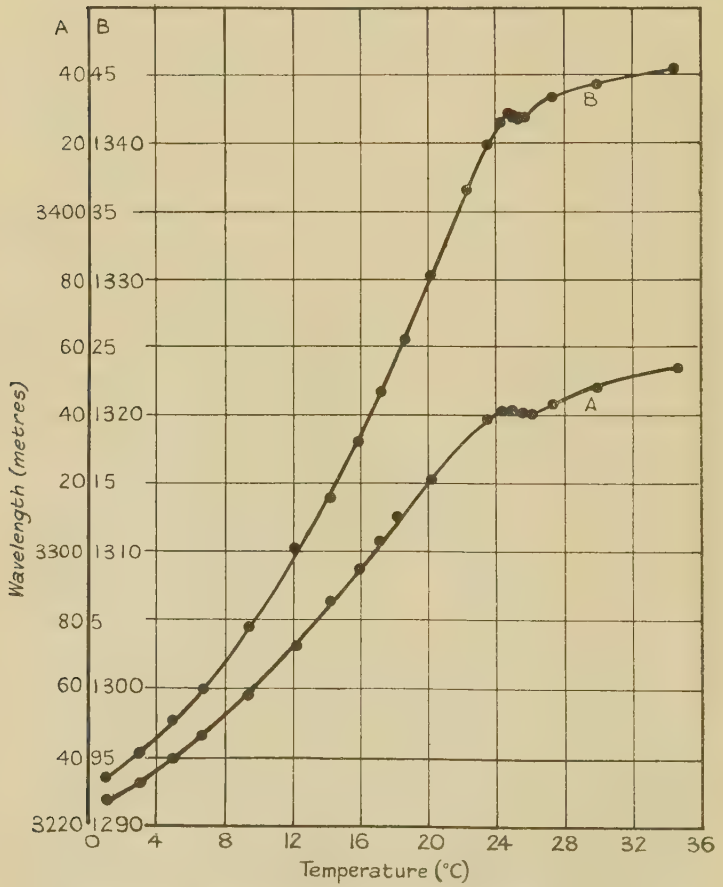


Resonant wave-length-temperature curves for resonator II. *a*.

shows that a discontinuity occurs also between −14° C. and −15° C. For resonator I. *a* the critical temperature is 23° C., this temperature being taken as the mean of the temperatures corresponding to the maximum and minimum at the discontinuity ; for resonator II. *a* the critical temperature,

similarly evaluated, is 23·5° C. Since similar changes occur at these temperatures in other physical properties of the crystal, it is clear that some fundamental change must occur in the crystal. One possible explanation is a change in

Fig. 6.



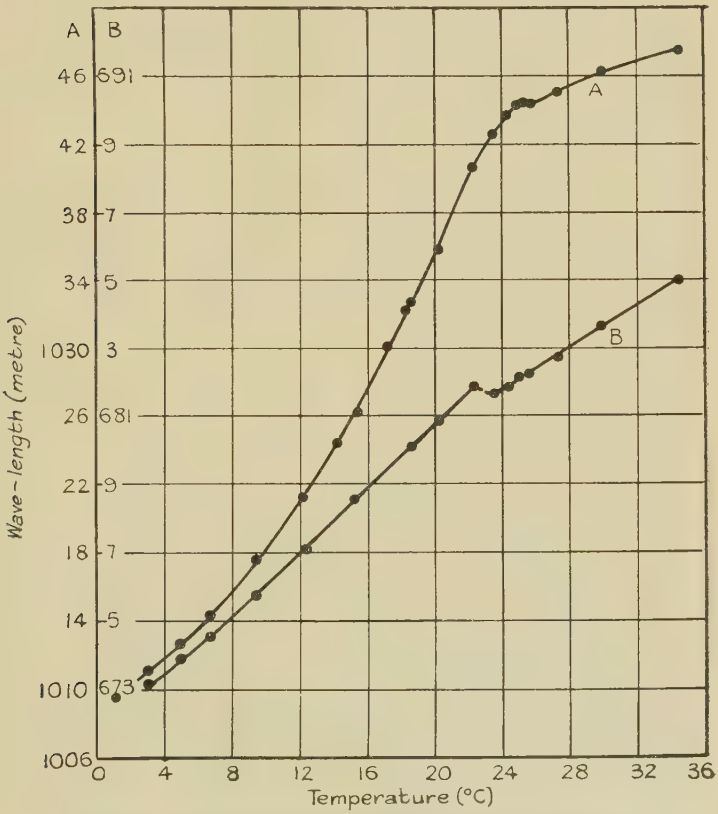
Resonant wave-length-temperature curves for resonator III. *a*.

crystal structure such as occurs in the analogous case of quartz; another possible explanation is that given by Schulwas-Sorokin [3a], in which the discontinuities in the various physical properties are attributed to a rotation at the critical temperatures of the bipoles constituting the crystal. In this connexion an analysis of the structure of

the Rochelle Salt crystal at different temperatures, by X-ray methods, would be of interest.

The temperature-coefficient of λ_e can be deduced from figs. 3 and 4. For resonator I. *a*, fundamental mode, it is $4{\cdot}8_5 \times 10^{-4}$ per °C.; for resonator II. *a*, fundamental mode, it is $9{\cdot}5_7 \times 10^{-4}$ per °C. The corresponding quantity for quartz is of the order of 5×10^{-6} per °C.

Fig. 7.



Resonant wave-length-temperature curves for resonator III. *a*.

The curves for resonator III. *a*, with one exception, are of a different type. The relation between λ_e and temperature is no longer linear, as was the case for resonators I. *a* and II. *a*. Other important differences are that the discontinuities in the curves are not so well defined, and that the critical temperature is in the neighbourhood of 25° C.

It is to be noted that most investigators have confined their attention to phenomena in the direction of the length of resonator II.*a*—the direction at 45° to the $O\bar{b}$ and O*c* axes. An examination of the temperature variation of the various properties in the direction of the length of resonator III.*a* would be of considerable interest in view of the results given here.

Finally, it may be stated that the large temperature coefficient of λ_e makes Rochelle Salt unsuitable for use in accurate work; to obtain the constancy of wave-length at present obtainable with quartz would require better temperature-control than is generally available. Where only moderate accuracy is required, Rochelle Salt resonators are useful, and can be easily made.

The author desires to express his sincere thanks to Prof. Gwilym Owen for his constant interest and encouragement; to Dr. W. Mandell, who undertook the tedious and difficult task of cutting and grinding the resonators; and to Prof. G. A. Schott, F.R.S., for checking the formulæ in the theoretical parts of the paper.

References.

(1) P. Vigoureux, 'Quartz Resonators and Oscillators' (H.M.S.O. 1931).
(2) W. Mandell, Proc. Roy. Soc. cxvi. p. 623 (1927).
(2*a*) R. M. Davies, 'Nature,' cxx. p. 332 (1927).
(3) J. Valasek, Phys. Rev. xix. p. 478 (1922).
(3*a*) R. D. Schulwas-Sorokin, *Zeit. für Phys.* lxxvii. p. 541 (1932).
(4) P. Vigoureux, *op. cit.* §§ 9, 10.
(5) Sir W. H. Bragg and R. E. Gibbs, Proc. Roy. Soc. cx. p. 443 (1926).
(6) W. G. Cady, Proc. Inst. Radio Engrs. x. p. 83 (1922).
(7) W. G. Cady, *loc. cit.*
(8) D. W. Dye, Proc. Phys. Soc. xxxviii. p. 399 (1926).
(9) P. Vigoureux, *op. cit.* §§ 33 and 37.
(10) D. W. Dye, *loc. cit.*; see also P. Vigoureux, *op. cit.* § 34 *et seq.*
(11) W. Voigt, 'Lehrbuch der Kristallphysik,' § 447.
(12) W. Voigt, *op. cit.* § 416.
(13) W. Voigt, *op. cit.* § 418.
(14) W. Voigt, *op. cit.* § 379.
(15) Lord Rayleigh, 'Theory of Sound,' vol. i. § 157 (1894).
(16) M. v. Laue, *Zeit. für Phys.* xxxiv. p. 347 (1925).
(17) E. Giebe and A. Scheibe, *Ann. der Phys.* ix. pp. 93 & 137 1931).
(18) W. Voigt, *op. cit.* § 280.
(19) W. Voigt, *op. cit.* § 289.
(20) P. Vigoureux, *op. cit.* § 23.
(21) W. Voigt, *op. cit.* § 392.
(22) J. Valasek, Phys. Rev. xx. p. 639 (1922).

VIII. *The Effect of the Earth's Magnetic Field on the Propagation of Short Wireless Waves.* *By* J. A. Ratcliffe, *M.A.*, *and* E. L. C. White, *B.A.**

[Plates II.–IV.]

1. *Introduction.*

IT has often been suggested that when wireless waves are transmitted through the ionosphere their behaviour will be influenced to a very large extent by the presence of the earth's magnetic field. A complete theory of the propagation of an electromagnetic wave through an ionized medium in the presence of a steady magnetic field has been worked out [1] [2] [3]. This magneto-ionic theory shows that, for a given ionization density and a given direction of the wave-normal relative to the magnetic field, two waves with characteristic polarizations and velocities may be propagated. As the characteristic polarizations and velocities depend both on the ionic density and on the direction of the wave-normal, it is clear that grave difficulties attend a theoretical consideration of the general case in which a wave is incident on a medium of changing ionization density in a direction inclined to the ionic gradient in the medium. A considerable simplification results, however, if we restrict ourselves to the case where the wave is incident on the medium along the direction of the ionic gradient, for in this case the wave-normal remains always along the same direction, and hence makes a constant angle with the magnetic field. These conditions are fulfilled when waves are received after being returned at vertical incidence by the ionosphere in which the ionization gradient is vertical, and by observing such vertically reflected waves it is possible to make an experimental test of some of the consequences of the magneto-ionic theory. It is the purpose of this paper to describe such an experimental test, and to discuss the bearing of the results obtained on our knowledge of the electrical structure of the upper atmosphere.

From a consideration of the magneto-ionic theory it appears that the behaviour of waves is fundamentally different according as their wave-length is greater or less

* Communicated by Professor E. V. Appleton, F.R.S.

than a critical wave-length which depends on the magnitude of the steady magnetic field. This critical wave-length, λ_c, is given by the expression

$$\lambda_c = \frac{2\pi mc^2}{He},$$

where m, e = mass and charge of the electron (e.s.u.),

H = magnitude of the earth's magnetic field (e.m.u.),

c = velocity of light in free space.

On the assumption that the carriers responsible for the upper atmospheric deviation are electrons, and that the value of the earth's field is 0·5 gauss, the critical wave-length (λ_c) is found to be approximately 200 m. It is therefore convenient to describe waves of length less than 200 m. as "short" and waves longer than 200 m. as "medium" and "long,". In this paper attention will be confined to a discussion of waves which are "short" in this sense.

The present series of experiments has been made possible by the development of a receiving system which, for waves incident in a vertical direction, responds only to circularly polarized waves with a certain sense of rotation *. The sense of rotation of the wave which is received can be altered instantaneously by pressing a key, so that any downcoming wave can be analyzed into two components circularly polarized in opposite directions. This apparatus may be called a circularly polarized receiver. By choosing a wave-length such that the two magneto-ionic components are circularly polarized on leaving the ionosphere it was therefore possible, by using this arrangement, to observe either component separately. Further, by combining this arrangement with the Breit and Tuve [4] echo method of observation, the effective height of the two components separately could be measured. The construction and action of the circularly polarized receiver will first be described, and then the results obtained will be discussed and compared with the predictions of the magneto-ionic theory.

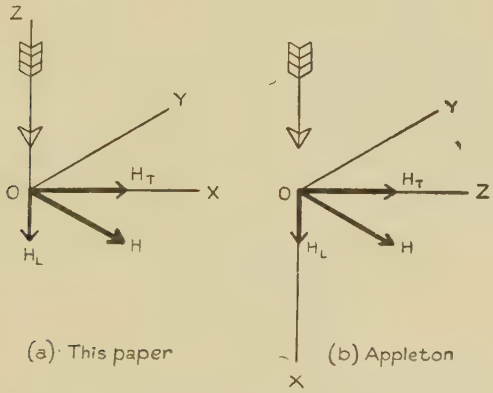
* A note on another type of apparatus with similar properties was published by Eckersley [5] during the course of the experiments here described.

As soon as the first results with the circularly polarized receiver had been obtained the authors were informed that Professor Appleton * had obtained similar results using a different method of observation. They wish to acknowledge the many helpful discussions which they have had with him in connexion with the results obtained.

2. *The Circularly Polarized Receiver.*

Suppose the downcoming wave is travelling along the axis of Z in the negative direction, as in fig. 1, and has

Fig. 1.



The axes used (*a*) in this paper,
(*b*) by Appleton in J. I. E. E. lxxi. p. 642 (1932).

an arbitrary type of polarization, and suppose its magnetic field has components $H_x e^{jpt}$ and $H_y e^{jpt}$. This wave may be thought of as the resultant of two circularly polarized waves

$$\{x = Pe^{jpt},\ y = jPe^{jpt}\} \quad \text{and} \quad \{x = Qe^{jpt},\ y = -jQe^{jpt}\},$$

with opposite senses of rotation. We then have

$$\left.\begin{aligned} &P = \tfrac{1}{2}(H_x - jH_y) \\ \text{and} \quad &Q = \tfrac{1}{2}(H_x + jH_y). \end{aligned}\right\} \quad . \quad . \quad . \quad . \quad . \quad . \quad (1)$$

The first of these waves will have a right-handed sense of rotation as seen by an observer looking along the

* The results of these experiments are referred to in Proc. Phys. Soc.

direction of travel of the wave, whereas the second wave will appear left-handed. The sense of polarization of a circularly (or elliptically) polarized wave will always be described in terms of the sense of rotation to an observer looking along the direction of travel of the wave. The first of the above-mentioned waves will

Fig. 2 (*a*).

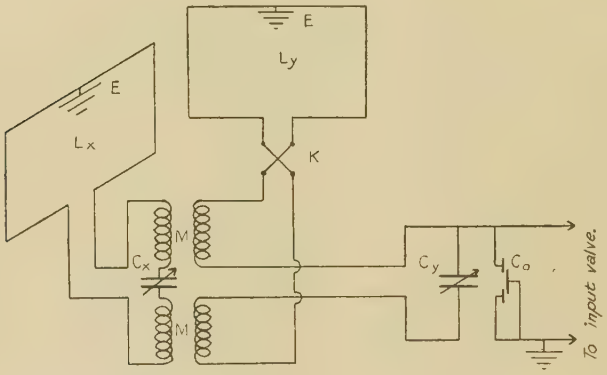
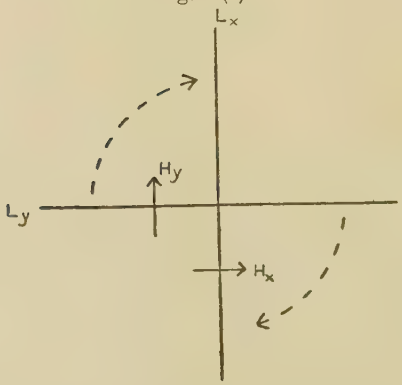


Fig. 2 (*b*).



therefore be called right-handed. Thus, if a receiver is constructed so that its response is a function of either $(H_x - jH_y)$ or $(H_x + jH_y)$, it will give a measure of the amplitude of either the right-handed or the left-handed component. The construction of such a receiver is accomplished in the following manner.

The aerial system consists of two similar centre-tapped looped aerials L_x and L_y at right angles to each other.

The loops are arranged with their planes vertical and with a common axis of symmetry, as shown in plan in fig. 2 (*b*). In fig. 2 (*a*) they are shown separately, in order to exhibit clearly their associated circuits. The loops are tuned separately by the condensers C_x and C_y, and are coupled by the split mutual inductance MM. The centre points of both loops are earthed. The input of an amplifier is connected from one side of the condenser C_y to the centre point of the loop L_y. The other loop circuit is left free. In order to balance out antenna effect in the loop connected to the amplifier the split balancing condenser C_0 is employed. The loop L_y is connected to its associated circuit through a reversing key K, consisting of a small telephone relay, which may be actuated by pressing a key at the observing point. The condenser C_x, the two halves of M, and the reversing relay are enclosed in a metal box placed immediately under the loops. Shielded leads are taken from this box to the loops and to the amplifier.

Suppose now that the downcoming wave has a magnetic field as shown in fig. 2 (*b*). Let i_x, i_y, and Z_x, Z_y respectively be the currents and vector impedances of the two circuits. Then we have

$$\left.\begin{aligned} i_x Z_x &= E_x - M\frac{di_y}{dt}, \\ i_y Z_y &= E_y - M\frac{di_x}{dt}, \end{aligned}\right\} \quad . \quad . \quad . \quad . \quad . \quad (2)$$

where

$$\left.\begin{aligned} E_x &= -A_x\frac{dH_x}{dt}, \\ E_y &= -A_y\frac{dH_y}{dt}, \end{aligned}\right\} \quad . \quad . \quad . \quad . \quad . \quad (3)$$

E_x and E_y being the e.m.f.'s induced when the magnetic fields of the wave link with the effective areas A_x and A_y of the loops. In the above equations the positive sense of the mutual inductance is taken as that which would exist if the two loops were turned parallel to each other, so that the positive direction of the magnetic field through them was the same, as indicated by the dotted arrows in fig. 2 (*b*). It is a simple matter to correlate the sense of the coupling in the coils M with the coupling

which would be produced if the loops were turned parallel to each other in this positive sense.

Returning to equations (2) and (3), and writing

$$i \equiv \mathrm{I}e^{jpt},$$

we have

$$\mathrm{I}_x\mathrm{Z}_x = -jp\mathrm{A}_x\mathrm{H}_x - jp\mathrm{MI}_y,$$

$$\mathrm{I}_y\mathrm{Z}_y = -jp\mathrm{A}_y\mathrm{H}_y - jp\mathrm{MI}_x,$$

which give

$$\mathrm{I}_x = \frac{-jp\mathrm{A}_x\mathrm{H}_x + \dfrac{jp\mathrm{M}}{\mathrm{Z}_y}jp\mathrm{A}_y\mathrm{H}_y}{\mathrm{Z}_x + \dfrac{p^2\mathrm{M}^2}{\mathrm{Z}_y}}.$$

If the circuit y is tuned so that Z_y is real and equal to R_y, and if the value of M is adjusted so that

$$p\mathrm{M} = \mathrm{R}_y\frac{\mathrm{A}_x}{\mathrm{A}_y},$$

then we find

$$\mathrm{I}_x = \frac{-jp\mathrm{A}_x}{\mathrm{Z}_x + \dfrac{p^2\mathrm{M}^2}{\mathrm{R}_y}}\{\mathrm{H}_x - j\mathrm{H}_y\}.$$

In virtue of equations (1) the response of this receiver is a function of P, the amplitude of the right-handed polarized wave. If the sense of M is reversed the receiver gives a response which is a function of $\{\mathrm{H}_x + j\mathrm{H}_y\}$, and hence of Q, the amplitude of the left-handed polarized wave.

The circuits are adjusted so as to satisfy the relations given above by observations made on the signal received *viâ* the ground ray. In the Breit and Tuve method of observation here used the ground signal is always separated (in time) from the echo signal, so that all the adjustments can be made even in the presence of a downcoming wave. If the signal is due to the direct wave with magnetic field H_0, and incident horizontally in a direction making an angle θ with the plane of the loop L_x, then

$$\mathrm{H}_x = \mathrm{H}_0\cos\theta \quad \text{and} \quad \mathrm{H}_y = \mathrm{H}_0\sin\theta,$$

so that if the conditions for circular polarization are satisfied we have

$$|\mathrm{I}_x| \propto \sqrt{\mathrm{H}_x^2 + \mathrm{H}_y^2} \propto \mathrm{H}_0.$$

This shows that under these conditions $|I_r|$ is independent of θ. The adjustment therefore consists in setting the mutual inductance M so that the ground signal is not varied as the loop system is rotated. In order that the sense of the received polarization should be reversed when the reversing key K is actuated it is necessary to ensure that there is no mutual inductance between the loops themselves. This is accomplished by adjusting the mutual position of the loops until the ground signal is not varied on pressing the key.

3. *Reflexion from the Appleton Region.*

The results of the magneto-ionic theory have been given by Appleton in previous papers [1]. On the assumption that the frequency of electronic collisions is small compared with frequency of the wave in the deviating regions of the ionosphere he showed that the refractive index (μ) was given by the expression

$$\mu^2 = 1 - \frac{2x}{2\left(1+\frac{x}{3}\right) - \frac{y_T^2}{\left(1-\frac{2x}{3}\right)} \mp \sqrt{\frac{y_T^4}{\left(1-\frac{2x}{3}\right)^2} + 4y_L^2}}, \quad (4)$$

$$x = \frac{4\pi N e^2}{mp^2}, \quad y_L = \frac{H_L e}{mcp}, \quad y_T = \frac{H_T e}{mcp};$$

where

m, e = mass, charge of ion (e.s.u.),

p = angular frequency of wave,

N = ion density in number per c.c.

H_L, H_T = components of earth's magnetic field along and perpendicular to the wave normal (e.m.u.),

c = velocity of light in free space.

The state of polarization of the wave is given by the expression

$$\frac{H_x}{H_y} = \frac{2iy_L x}{\frac{y_T^2}{\left(1-\frac{2x}{3}\right)} \pm \sqrt{\frac{y_T^4}{\left(1-\frac{2x}{3}\right)^2} + 4y_L^2}}, \quad (5)$$

where the wave is supposed to be travelling as shown in fig. 1. Equation (5) differs from Appleton's equation (37)

in two respects: first, a different set of axes is here taken, this being found much more convenient for a description of the practical arrangement of section 2 (the present axes are compared with those used by Appleton in fig. 1); secondly, the quantity $y_L\left(=\frac{H_L e}{mcp}\right)$ is here taken to be positive for positive values of the charge e. Appleton takes y to be positive for *negative* charges, so that a similar expression deduced from his equation (37) would have the opposite sign.

If we consider a wave of length 100 m. travelling vertically upwards into the ionosphere, and the earth's magnetic field be taken as 0·5 gauss directed downwards at an angle of 23° to the vertical (as is the case in England) then the way in which the refractive index and the state of polarization vary with the ionic density is represented in the curves of fig. 3. From this we see that the ordinary wave will penetrate to a height where the ionic density is given by the critical value

$$N_0=\frac{3}{2}\frac{mp^2}{4\pi e^2},$$

which reduces the refractive index for the ordinary wave to zero. The extraordinary wave will not penetrate to so great a height, but will be turned back when it encounters an ionic density

$$N_e=\frac{3}{2}\frac{mp^2}{4\pi e^2}(1-y).$$

where

$$y^2=y_L^2+y_T^2.$$

If, further, we assume that the ions responsible for the deviation have a negative charge, then equation (5) shows that on leaving the ionosphere (limiting case of $N=0$) the ordinary wave will be polarized with a left-handed sense of rotation. The curve of fig. 3(*b*) shows that the wave is almost exactly circularly polarized, the ratio of axes being 0·97. On leaving the ionosphere the extraordinary wave will have a right-handed direction of polarization. This theory would predict that if we send a plane-polarized wave vertically up into the ionosphere it should be returned vertically downwards, split into two components approximately circularly polarized with opposite senses of rotation. The left-handed component would have travelled further than the right-handed.

Appleton and Builder [5] have pointed out that when waves are returned at vertical incidence from the Appleton or F region of the ionosphere (250 to 300 km. equivalent height) two different times of travel are

Fig. 3 (*a*).

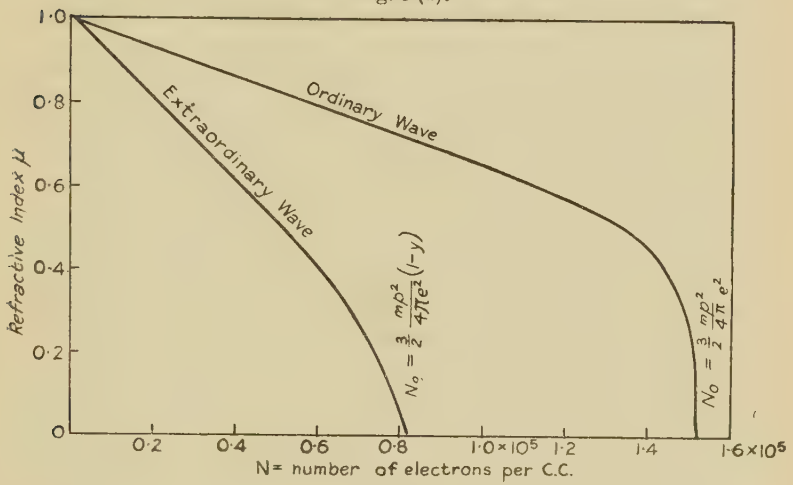
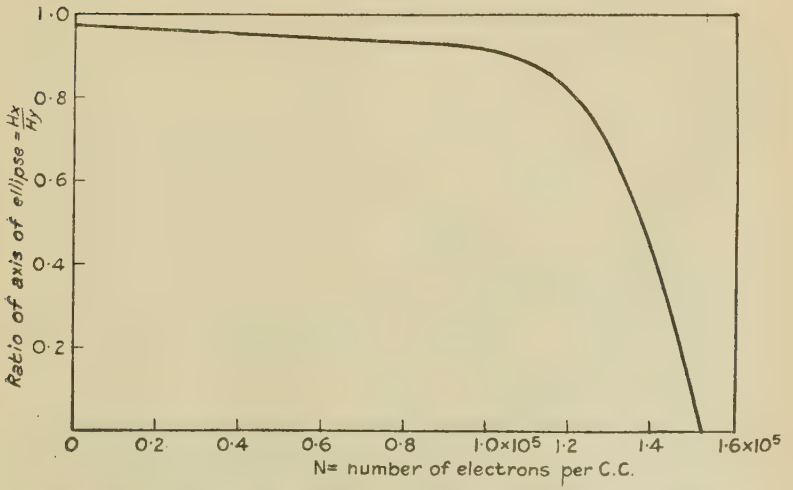


Fig. 3 (*b*).



recorded when the ionization has decreased sufficiently to bring the ionic density near the critical value N_0. If the magneto-ionic theory is correct the uppermost echo would be expected to be circularly polarized with left-

handed rotation and the lower to be circularly polarized with right-handed rotation. This was the first point to be investigated using the circularly polarized receiver, and the type of result obtained is shown in fig. 4 (Pl. II.), which is reproduced from a photograph of the echo patterns observed on a cathode ray oscillograph screen. (*a*) shows the doubled echo obtained using an ordinary loop aerial, (*b*) the upper component alone is received with the circularly polarized receiver set so as to receive only left-handed circularly polarized waves (which setting we will denote by L), and (*c*) the lower component alone is received with the receiver set so as to receive only right-handed circularly polarized waves (condition R). This type of behaviour is always observed when a split echo occurs due to the ionization in the F region approaching the critical value N_0. It is clear that the prediction of the theory is completely verified.

The method of automatic recording elsewhere described [6] has been applied to the circularly polarized receiver, and the type of record obtained is shown in fig. 5 (Pl. II.). The receiver is arranged so that it is switched automatically from condition R to condition L at intervals of about five minutes. When the receiver is set to condition R a small lamp is lit which is recorded on the photographic film as a short line at the top of the record (as shown at *a* in the figure). We see that the uppermost component is polarized left-handed, while the lower one is right-handed. Records of this kind indicate the presence of the splitting even when the two components overlap, so that ordinarily they would not be resolved. Thus between 2000 and 2130 it is clear that the effective height of the left-handed component is greater than that of the right-handed, and yet on a simple aerial we should only observe a single rather broad echo. The circularly polarized receiver thus increases the resolving power of the apparatus as regards the magnetic splitting.

At the upper edge of fig. 5 (Pl. II.), between 2300 and 2400, a trace is seen which is due to a signal twice reflected from the Appleton region. It is clear that the polarization of the two components is the same as in the single reflexion but that their separation is twice as great *.

* The polarization of multiple echoes is also shown in fig. 8 (Pl. III.) 2200 to 2400 G.M.T.

This double separation, and the absence of more than two echoes in the twice reflected wave, would be difficult to explain on the hypothesis of a double stratification in the ionization, but is just what would be expected on the magneto-ionic theory of the doubling. A case where a different type of doubling occurs will be mentioned later.

It does not always happen that the echo is split into two sharp and well-defined components. Often instead it merely becomes broadened and may have a total width corresponding to an extension in effective height of as much as 200 km. Under these conditions the broadened echo often seems to consist of a series of separate peaks which fade rapidly in and out, giving a characteristic "boiling" appearance. It is usually found that, with the circularly polarized receiver set either to R or to L, a broadened echo is received, but when it is set to R the echo extends to lower effective heights and as a rule is very much sharper. Fig. 6 (Pl. II.) shows a photograph taken with the automatic recorder under conditions where the upper portion of the echo is partly right- and partly left-handed. In general the short delay edge of the echo has right-handed polarization, while the longer delay portion may have both left- and right-handed polarization.

An explanation of this phenomenon is probably connected with the fact that at the point N_e in fig. 3 (*a*), where the extraordinary wave is returned, the state of polarization has not changed very much (see fig. 3 (*b*)), and all the right-handed component in the *upgoing* wave will be reflected at this point. Near the point N_0, however, the state of polarization is changing very rapidly, and it may well be that some right-handed component is produced as the left-handed wave passes through this region. Thus the uppermost portion of the echo, reflected from near N_0, may contain both left- and right-handed components. The two components will have different group velocities, and if a continual interchange of types of polarization is taking place a broadened echo, showing both polarizations, may result.

4. *The Transition from* E *Region to* F *Region.*

The previous section has dealt only with reflexion from the Appleton or F region. In this section those

phenomena will be discussed which may be observed during the change-over from reflexion at the E region to reflexion at the F region. It is well known that during the day an E region echo with an effective height of about 120 km. is observed alone (accompanied possibly by its multiples), and that towards sunset an F region echo appears at a height of about 250 km., and the two echoes exist together for some time. Finally the E echo disappears and the F echo is left alone. Some writers have considered that there is a difficulty in explaining the simultaneous appearance of an E echo and an F echo [7]. If, however, it is supposed that the E region reflexion is due to the presence of electrons, then the sequence of events near sunset may be simply explained on the magneto-ionic theory as follows. Due to recombination the electron density in the E region drops to the critical value N_0 where the ordinary wave can no longer be returned, so that this wave penetrates the E region and is reflected from the F region. The extraordinary wave, however, is still returned from the E region, so that we have echoes from both regions existing simultaneously. If this explanation is the correct one then the extraordinary wave returned from the E region should be circularly polarized with a right-handed sense of rotation, whereas the ordinary echo returned from the F region should have a left-handed rotation. This can be tested with the circularly polarized receiver. Fig. 7 (Pl. III.) shows the kind of result obtained during the transition period when E and F echoes are observed simultaneously, and it completely bears out the explanation suggested above.

Although the above-mentioned kind of behaviour is found to occur during a large portion of the transition period it does not seem to be the only mechanism which gives rise to E and F echoes simultaneously. Sometimes it is found that each echo contains both right- and left-handed components. Under these circumstances it is probable that the E echo is due to partial reflexion at a sharp gradient of ionization, whereas the F echo is due to the remaining energy which has penetrated the E region and been returned from the F region. The conditions for such partial reflexion have been discussed by several writers [9]. On the whole it is found that simultaneous

reflexion under these conditions is much less common than under the first-mentioned conditions*.

In the case of the E region the penetration by the two components separately is not so obvious as it is in the case of the F region. In the latter case the echo splits and each component rises in turn to a great height and then disappears, so that the time interval between the two disappearances is definite and clear cut even without the use of a polarized receiver. In the case of the E region, however, there is no obvious splitting of the echo, and ordinary methods of observation give no clue as to whether or not one of the component waves has penetrated this region. It is not until observations are made with the circularly polarized receiver that the penetration by the two components at different times is demonstrated. The time at which the ordinary wave ceases to be reflected from the E region gives the time at which the ionic density reaches the value N_0, whereas the disappearance of the echo as a whole gives the time at which the ionic density reaches the value N_e. From the difference of these two times the recombination coefficient for the E region can be calculated.

The fact that this penetration occurs at different times has an important bearing on the question of the nature of the charged particles which compose the E region. Thus in order to account for the great difference in the behaviour of the ordinary and extraordinary waves it is necessary to suppose that the wave-length used is not far removed from the critical wave-length for the E region. This leads to the conclusion that the E region reflexion is largely due to electrons; if it were due to ions the critical wave-length $\left(\frac{2\pi mc^2}{He}\right)$ would be greater than 500,000 m.

The magneto-ionic explanation of the simultaneous reflexion from the E and F regions clears up another point which is otherwise difficult to explain. It is often found that the reflected waves from both regions are so strong that several multiple reflexions from the E region and from the F region occur simultaneously. Thus

* It is, however, noticeable that with the "abnormal" E reflexion, which sometimes occurs in the middle of the night, this type of partial reflexion is very common. This may be due to the fact that the E region is then very sharp, or to the fact that absorption is small so that weak reflexions are noticed.

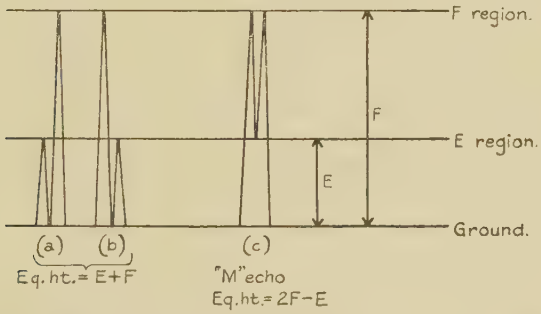
between 2300 G.M.T. and 2400 G.M.T. in the record of fig. 8 (Pl. III.) echoes at heights of 120, 240, 360 km. are due to waves reflected rom the E region once, twice, and three times respectively, and echoes at 320, 640 km. are due to waves reflected once and twice respectively from the F region. Now, although these multiple reflexions from each region separately are so common, it is found that echoes with an effective height corresponding to a reflexion from E followed by a reflexion from F (E+F echoes) or a reflexion from F followed by a reflexion from E (F+E echoes) are extremely uncommon. (Actually we have *never* observed echoes of these types, but this may be due to insufficient signal strength at the transmitter.) On a partial reflexion theory the absence of these echoes is difficult to explain, but on the magneto-ionic theory a simple explanation may be given as follows.

First, it is necessary to consider the change of polarization which occurs when a wave is reflected from the ground. For normal incidence, any elliptically polarized wave is reflected from the ground as a similar ellipse, whatever the electrical properties of the ground. The absolute direction of rotation of the ellipse remains unchanged, so that the sense of polarization of the wave as seen by an observer looking along the direction of travel is reversed after reflexion. Since the reflected wave is now incident on the ionosphere with a reversed direction of travel and with a reversed sense of polarization, its velocity at any point is the same as it was on its downward journey, so that the behaviour of any given wave is described by the same curve in fig. 3 whether it is travelling upwards after reflexion or downwards before reflexion at the ground. In view of this simplification there is no need to consider in detail the direction of travel of the wave and the sense of rotation relative to that direction; it is sufficient to state the *absolute* sense of rotation, and then the behaviour of the wave is known whether it is travelling upwards or downwards. Upgoing and downgoing waves will therefore be described by their sense of rotation as seen by an observer looking towards the ground; this will agree with the previous nomenclature for a wave travelling downwards. It has been shown that during the period of transition between the E and F regions the E region is capable of reflecting waves which reach it with right-handed polarization,

but that it transmits those reaching it with left-handed polarization. A wave which has once been returned from the E region will therefore have a right-handed polarization, and if it is reflected from the ground it must be returned a second time from the E region; it cannot penetrate to the F region, because the E region will not let it through. Similarly a wave once transmitted by the E region and reflected by the F region will have a left-handed polarization and will always pass through the E region after reflexion from the ground, and can only be reflected a second time when it reaches the F region. Thus echoes E, 2E, 3E; F, 2F, 3F, etc. may occur quite strongly, whereas E+F or F+E are completely absent.

Although echoes of the type shown in fig. 9, (a) and (b)

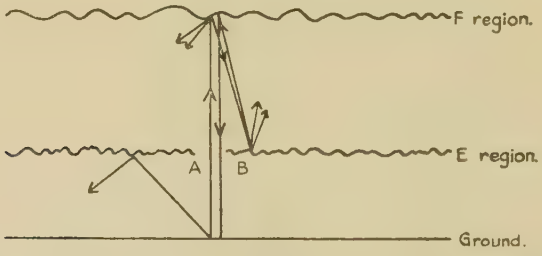
Fig. 9



are not observed, yet it is quite common to observe an echo with an effective height corresponding to a reflexion of the type shown at (*c*). We will denote this type of echo by M. It occurs most frequently just after the E region has finally been penetrated by the extraordinary wave, or just before the E echo reappears, due to increase of ionization in the E region. By using the circularly polarized receiver it is found that the M echo is circularly polarized with right-handed sense of rotation. Fig. 10 (Pl. IV.) is a reproduction of a photograph which illustrates this point. The M echo is also clearly visible, with right-handed polarization, between 2100 and 2400 G.M.T. on the record of fig. 8 (Pl. III.). A possible explanation of it is as follows:—When the right-handed (extraordinary) wave has just penetrated the E region for vertical incidence it will still be incapable of penetrating

it at oblique incidence. It is as if a small hole had been made in the E region vertically above the transmitter through which the waves could pass, as indicated in fig. 11. If it is supposed that these waves may be reflected somewhat diffusely at the F region, some portion of them, on their downward travel, may be incident on the E region, obliquely as at B, so that they cannot be transmitted by it. They will therefore be returned to the F region, possibly again by a somewhat diffuse reflexion, and, after a further diffuse reflexion at the F region, they may pass vertically downwards through the E region. The suggested course of the waves is shown in the figure. It is clear that, for the reflexion at oblique incidence at B to be possible, the ionization density must be nearly critical, and this is only the case

Fig. 11.



under the assumed conditions for the right-handed wave. This agrees with the experimental result that the M echo has right-handed polarization and that it occurs when the E region ionization is nearly critical.

Another phenomenon is sometimes observed which lends support to the present theory of the change over from E to F reflexion. Just after the E echo has disappeared completely it is occasionally found that the F echo is doubled, with a separation corresponding to an equivalent height of about 20 to 40 km. *. Examination with the circularly polarized receiver shows that the *uppermost* component of the doublet has right-handed polarization and the *lower* one is left-handed. This is exactly the opposite of what is observed in the case of the splitting of the F echo due to electron limitation in the F region. Fig. 12 (Pl. IV.) shows the phenomenon

* From the magnitude of this separation it seems possible to make deductions about the density of ionization between the E and F regions.

as photographed on an automatic record. Between 1800 and 1930 G.M.T. the "reversed splitting," with R uppermost, is seen to occur, whereas after 2100 G.M.T. the "normal splitting," with L uppermost, is observed.

The explanation of this effect seems to be as follows:— Just after the E region echo disappears entirely the maximum density of the E region is just less than the critical value for the extraordinary wave, and reference to fig. 4 (Pl. II.) shows that, under these conditions, the phase velocity of this wave in the E region will be large. Under the same conditions the phase velocity of the ordinary wave will not be far different from the velocity in free space. Now, for these wave-lengths, when the phase velocity is large the group velocity is small, so that any wave which traverses a region where the phase velocity is very large will be expected to show an increased group time. It is now apparent that the extraordinary wave suffers an extra group delay in passing through the E region with the ionic density nearly critical, whereas the ordinary wave suffers no such retardation. As recombination in the E region progresses the ionization ceases to be nearly critical for the extraordinary wave, so that the corresponding echo ceases to suffer the extra group delay, and it decreases in effective height, and finally coincides with the ordinary echo as seen at 1930 in the record of fig. 12 (Pl. IV.)*.

5. *The* E *Region.*

It has been pointed out in the previous section that the E region echo is probably due to electrons, and attention has been drawn to the fact that this echo does not exhibit appreciable doubling like that found with the F region echo. Observations on the echo from the "abnormal E region"†, which sometimes occurs in the middle of the night, have occasionally shown a double echo. It appears, however, that the doubling here is sometimes different from that which characterizes the F region. The first point of difference is that the

* The effect of E region retardation in increasing the effective height of the F region for *both* components is shown at 2430 G.M.T. in fig. 8 (Pl. III.).

† The name "abnormal E region" is here given to the increase of ionization which is sometimes found in the E region during the hours of darkness. *Vide* Appleton and Naismith, Proc. Roy. Soc. cxxxvii. A, p. 36 (1932).

doublet separation is *not* increased in the higher order reflexions; it seems to remain constant or even to show a slight decrease. This behaviour is quite different from that of the F doublet described in section 3 and illustrated in fig. 5 (Pl. II.). A second point of difference is that both components of the double E echo are polarized in the same sense. The state of polarization may change rapidly, but it is found that at a given instant both components have the same type of polarization. In view of the two experimental facts it is probable that the doubling of the "abnormal E region" echo is not always of magneto-ionic origin, but is due to a double stratification of the region.

6. *Absorption Effects.*

Experimental Results.

By means of the circularly polarized receiver observations may be made of the relative intensities of the two downcoming components even when they have not different delay times; it is merely necessary to compare the amplitudes as the reversing key is switched over. The effects are somewhat variable and are much influenced by natural fading, but the following broad deductions may be drawn from the observations :—

(1) At night for the wave-lengths we have studied (60–180 m.) both right- and left-handed components are roughly equal in magnitude except under conditions where one of them is about to disappear, due to electron limitation.

(2) In the daytime the left-handed component is of greater intensity than the right-handed. Both are much weaker than at night.

Both the above observations refer to reflexion from either region. The case of reflexion from the E region during the day is illustrated in fig. 13 (Pl. IV.), where it is seen that only the left-handed component was received with sufficient intensity to be recorded.

Theory.

Several lines of argument lead to the conclusion that for these wave-lengths most of the absorption of a wave must occur either in or just below the E region. It is fairly certain that in this region the frequency of the electron collisions is less than the frequency of the wave.

Under these conditions Appleton's theory shows that the absorption coefficient κ is given by the expression

$$\frac{\kappa c}{p} = \frac{(1-\mu^2)^2}{2\mu}\left\{\frac{z}{x}\left(1+\frac{y_T^2}{2\left(1-\frac{2x}{3}\right)^2}\right.\right.$$
$$\left.\left.\mp\frac{y_T^4}{2\left(1-\frac{2x}{3}\right)^3\sqrt{y_T^4+4y_L^2\left(1-\frac{2x}{3}\right)^2}}\right)\right\}. \quad (6)$$

The symbols here have the same meanings as in section 3, and $z=\frac{\nu}{p}$, where ν is the frequency of the electron collisions.

For a wave which is being reflected from the F region the ionization density in the E region is usually not near the critical value, so μ is approximately equal to unity. If we write $\mu(=)1$ in the denominator of equation (6), and then substitute for $(1-\mu^2)$ from equation (4), we obtain

$$\frac{\kappa c}{p} = \frac{z}{2x}\left\{\frac{2x}{2\left(1+\frac{x}{3}\right)-\frac{y_T^2}{1-\frac{2x}{3}}\mp\sqrt{\left[\frac{y_T^4}{\left(1-\frac{2x}{3}\right)^2}+4y_L^2\right]}}\right\}^2$$
$$\left\{1+\frac{y_T^2}{2\left(1-\frac{2x}{3}\right)^2}\mp\frac{y_T^4}{2\left(1-\frac{2x}{3}\right)^3\sqrt{\left[y_T^4+y_L^2\left(1-\frac{2x}{3}\right)^2\right]}}\right\}. \quad (7)$$

In both the factors in this expression the lower sign refers to the ordinary wave. Now for 100 m. waves coming vertically downwards at an angle of 30° to the magnetic field $y_T^2=0{\cdot}06$. It is then seen that the upper sign gives the greater value of the first factor, whereas it gives the smaller value of the second factor. Thus it is impossible, without further calculation, to say which of the factors will predominate, and whether κ will be greater for the lower or the upper sign. If, however, we put in numerical values we find that for our case the difference in the values for the two signs is determined by the first factor, the variation in the second factor being negligible. It is then clear that the absorption is greater for the wave corresponding to the upper sign, *i.e.*, for the extraordinary wave. This is in agreement with the observations men-

tioned above. These observations and the theory here given yield results directly opposed to those mentioned by Eckersley [8]. The discrepancy is not yet understood *.

From equation (7) it may be seen that the difference in the absorption of the left- and right-handed components is more marked the longer the wave-length (within the present range of "short" wave-lengths). In using various wave-lengths between 60 and 180 m. this is very noticeable when the polarized receiver is employed. It may be shown that the effects of differential absorption are much more marked when the overall absorption is great (*i. e.*, in the day). The observations recorded above are in agreement with this deduction.

We wish to express our thanks to Prof. F. J. M. Stratton for providing us with an ideal site for the transmitting apparatus at the Solar Physics Observatory, Cambridge, and for his readiness at all times to provide assistance for our experiments.

Summary.

The paper describes the construction and use of a circularly polarized receiver which makes it possible to pick out the right- and left-handed circularly polarized components for a wave incident vertically.

The polarized receiver is used to test several deductions from the magneto-ionic theory, for a wave-length less than the critical magneto-ionic wave-length of 214 m. Magneto-ionic effects are found in both the upper (F) region and the lower (E) region; this points to the fact that electrons are responsible for the deviation of the waves in both these regions. The occurrence of simultaneous reflexions from the E and the F regions is investigated, and the polarization of daytime signals is discussed.

References.

(1) Appleton, Journ. I. E. E. lxxi. p. 642 (1932); U.R.S.I. Papers (Washington, 1927).
(2) Goldstein, Proc. Roy. Soc. cxxi. A, p. 260 (1928).
(3) Hartree, Proc. Camb. Phil. Soc. xxvii. p. 143 (1931).
(4) Breit & Tuve, Phys. Rev. xxviii. p. 554 (1928).
(5) Appleton & Builder, Proc. Phys. Soc. xliv. p. 76 (1932).
(6) White, 'Nature,' cxxix. p. 579 (1932).
(7) Schafer & Goodall, Proc. I.R.E. xx. p. 1139 (1932).
(8) Eckersley, 'Nature,' cxxx. p. 398 (1932).
(9) Hartree, Proc. Roy. Soc. cxxxi. A, p. 427 (1931).

* In this connexion reference should be made to a note by Appleton and Ratcliffe, 'Nature,' Sept. 24th, 1932.

IX. *Temperature Stresses in Walls and Flat Slabs.* *By* Norman S. Boulton, *M.Sc.*, *Assoc.M.Inst.C.E.*, *Armstrong College*, *Newcastle-upon-Tyne* *.

Synopsis.

The following investigation has for its object the determination of the stresses induced in a wall or flat slab of steel, concrete, or other fairly homogeneous elastic material, due to the heating of one face above the temperature of the other. Attention is mainly confined to a wall or slab the edges of which are free from external restraint. It will appear, in what follows, that the stresses considered only arise when the temperature gradient through the wall is non-uniform, which condition implies that the temperature at any point is varying with the time.

In § 1 equations are deduced for the stresses and displacements in the slab, without assuming a particular law connecting the temperature with the time. In § 2 an equation is derived which, in § 3, is shown to represent approximately the variable temperature distribution in a flat slab due to the heating of one face by the sun's rays. Expressions for the integrals in the stress equations of § 1, and for the maximum stress at the heated face, are also deduced in § 2 from this equation. This application of the general solution in § 1 is of some interest and importance as indicating whether temperature stresses may be partly responsible for the disintegration of the faces of walls exposed to sunlight. Numerical values of the stresses in this case, and the conclusions arising therefrom, are considered in § 4. The limitations of the solution near the edges of the slab are investigated in § 5.

---

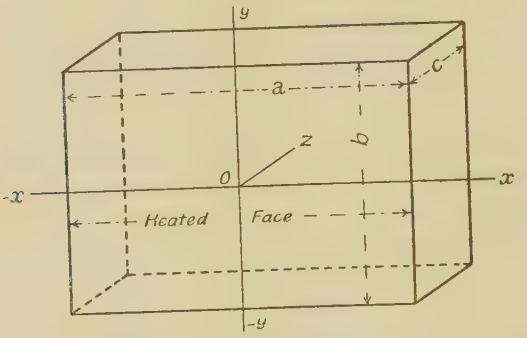
§ 1. *Stress and Displacement Equations.*

USING rectangular coordinates, the origin of which is taken at the centre of the heated face, let the slab be bounded by the planes for which $x = \pm a/2$, $y = \pm b/2$, $z = 0$, and $z = c$ (fig. 1). In what follows, the first four planes will be referred to as the *edges*, and the last two as the *faces*, of the slab.

* Communicated by the Author.

We assume that the slab is initially free from stress, that due to its own weight being neglected, that the coefficient of expansion and the elastic constants of the material do not vary appreciably over the range of temperatures considered; and that the elastic limits of the material are not exceeded. We further suppose that the face $z=0$ is heated uniformly so that heat flows in the z-direction only.

Fig. 1.



When differences of temperature exist in a body, the stress-strain equations are*:

$$\left.\begin{aligned}\frac{\partial u}{\partial x}&=\frac{1}{\mathrm{E}}\left\{\widehat{xx}-\sigma(\widehat{yy}+\widehat{zz})\right\}+\alpha\theta,\\ \frac{\partial v}{\partial y}&=\frac{1}{\mathrm{E}}\left\{\widehat{yy}-\sigma(\widehat{zz}+\widehat{xx})\right\}+\alpha\theta,\\ \frac{\partial w}{\partial z}&=\frac{1}{\mathrm{E}}\left\{\widehat{zz}-\sigma(\widehat{xx}+\widehat{yy})\right\}+\alpha\theta,\end{aligned}\right\}\quad\ldots\quad(1.1)$$

$$\left.\begin{aligned}\frac{\partial u}{\partial y}+\frac{\partial v}{\partial x}&=\frac{1}{\mu}.\widehat{xy},\\ \frac{\partial v}{\partial z}+\frac{\partial w}{\partial y}&=\frac{1}{\mu}.\widehat{yz},\\ \frac{\partial w}{\partial x}+\frac{\partial u}{\partial z}&=\frac{1}{\mu}.\widehat{zx},\end{aligned}\right\}\quad\ldots\ldots\quad(1.2)$$

in which θ is the excess temperature at any point (x, y, z) above the initial uniform temperature of the slab, α is the linear coefficient of expansion, E is Young's modulus, σ is

* Love, 'Mathematical Theory of Elasticity,' 4th ed. pp. 125, 126, 108.

Poisson's ratio, and μ is the modulus of rigidity. Also the stress equations are, when there are no body forces,

$$\left.\begin{aligned} \frac{\partial \widehat{xx}}{\partial x} + \frac{\partial \widehat{xy}}{\partial y} + \frac{\partial \widehat{zx}}{\partial z} &= 0, \\ \frac{\partial \widehat{xy}}{\partial x} + \frac{\partial \widehat{yy}}{\partial y} + \frac{\partial \widehat{yz}}{\partial z} &= 0, \\ \frac{\partial \widehat{zx}}{\partial x} + \frac{\partial \widehat{yz}}{\partial y} + \frac{\partial \widehat{zz}}{\partial z} &= 0. \end{aligned}\right\} \quad . \quad . \quad . \quad (1.3)$$

Let us now tentatively assume that the stresses are constant for all points of any plane parallel to the faces of the slab, *i. e.*, that $\widehat{xx} = \widehat{yy} = p$ (say), p being independent of x and y. Further assume that the shear stresses and the stress $\widehat{zz}$ are everywhere zero. Equations (1.3) are clearly satisfied by this state of stress, which can therefore exist provided that suitable boundary stresses are applied to the edges of the slab, and that the displacements u, v, w can be found so as to satisfy equations (1.1) and (1.2).

On the above assumptions, equations (1.1) now become

$$\frac{\partial u}{\partial x} = \frac{\partial v}{\partial y} = \frac{1-\sigma}{\mathrm{E}} \cdot p + \alpha\theta, \quad . \quad . \quad . \quad . \quad (1.4)$$

$$\frac{\partial w}{\partial z} = -\frac{2\sigma}{\mathrm{E}} \cdot p + \alpha\theta, \quad . \quad . \quad . \quad . \quad (1.5)$$

and (1.2) become

$$\frac{\partial u}{\partial y} + \frac{\partial v}{\partial x} = 0, \quad . \quad . \quad . \quad . \quad . \quad . \quad (1.6)$$

$$\frac{\partial v}{\partial z} + \frac{\partial w}{\partial y} = 0, \quad . \quad . \quad . \quad . \quad . \quad . \quad (1.7)$$

$$\frac{\partial w}{\partial x} + \frac{\partial u}{\partial z} = 0. \quad . \quad . \quad . \quad . \quad . \quad . \quad (1.8)$$

To find the stresses and displacements, integrate (1.5) and we have

$$w = -\frac{2\sigma}{\mathrm{E}} \int p\, dz + \alpha \int \theta\, dz + w_0, \quad . \quad . \quad . \quad (1.9)$$

where w_0 is a function of x and y.

Also (1.8) and (1.7) give

$$\frac{\partial u}{\partial z} = -\frac{\partial w}{\partial x}, \quad \frac{\partial v}{\partial z} = -\frac{\partial w}{\partial y},$$

which with (1.9) become

$$\frac{\partial u}{\partial z} = -\frac{\partial w_0}{\partial x}, \quad \frac{\partial v}{\partial z} = -\frac{\partial w_0}{\partial y}, \quad . \quad . \quad (1.10)$$

since the two integrals in (1.9) are independent of x and y.

Integrating (1.10), we must have

$$u = -z\frac{\partial w_0}{\partial x} + u_0, \quad . \quad . \quad . \quad . \quad . \quad (1.11)$$

$$v = -z\frac{\partial w_0}{\partial y} + v_0, \quad . \quad . \quad . \quad . \quad . \quad (1.12)$$

where u_0 and v_0 are functions of x and y. Now since the right-hand side of (1.4) is independent of x and y, differentiating the terms independent of z in (1.11) and (1.12), we must have

$$\frac{\partial u_0}{\partial x} = \frac{\partial v_0}{\partial y} = \mathrm{B}, \quad . \quad . \quad . \quad . \quad . \quad (1.13)$$

where B is a constant; and differentiating the terms containing z

$$\frac{\partial^2 w_0}{\partial x^2} = \frac{\partial^2 w_0}{\partial y^2} = -\mathrm{A}, \quad . \quad . \quad . \quad . \quad (1.14)$$

where A is a constant, w_0 being independent of z.

Again (1.6) with (1.11) and (1.12), gives

$$\frac{\partial u_0}{\partial y} + \frac{\partial v_0}{\partial x} = 0, \quad . \quad . \quad . \quad . \quad . \quad (1.15)$$

and

$$\frac{\partial^2 w_0}{\partial x . \partial y} = 0. \quad . \quad . \quad . \quad . \quad . \quad (1.16)$$

Now (1.13) shows that

$$u_0 = \mathrm{B}x + \mathrm{F}_1(y),$$

$$v_0 = \mathrm{B}y + \mathrm{F}_2(x),$$

where F_1 is a function of y, and F_2 is a function of x; and from (1.15) we must have

$$\frac{\partial \mathrm{F}_1(y)}{\partial y} + \frac{\partial \mathrm{F}_2(x)}{\partial x} = 0,$$

and this equation requires that

$$\frac{\partial \mathrm{F}_1(y)}{\partial y} \quad \text{and} \quad \frac{\partial \mathrm{F}_2(x)}{\partial x}$$

should be constants $= c''$ and $-c''$, say.

Hence

$$F_1(y)=c''y+a', \qquad F_2(x)=-c''x+b',$$

where a' and b' are constants, and

$$\left.\begin{aligned} u_0&=Bx+c''y+a',\\ v_0&=By-c''x+b', \end{aligned}\right\} \quad . \quad . \quad . \quad . \quad (1.17)$$

Now the only solution of (1.14) consistent with (1.16) is

$$w_0=-\frac{Ax^2}{2}-\frac{Ay^2}{2}+a''x+b''y+c', \quad . \quad . \quad (1.18)$$

where a'', b'', and c' are constants.

Differentiating (1.18),

$$-\frac{\partial w_0}{\partial x}=Ax-a''. \quad . \quad . \quad . \quad . \quad (1.19)$$

Hence (1.11), with (1.17) and (1.19), gives

$$u=z(Ax-a'')+Bx+c''y+a', \quad . \quad . \quad (1.20)$$

and therefore

$$\frac{\partial u}{\partial x}=Az+B.$$

Substituting for $\frac{\partial u}{\partial x}$ in (1.4),

$$Az+B=\frac{1-\sigma}{E}p+\alpha\theta,$$

and writing $A=\alpha M$, $B=\alpha N$, we have, for the stress in the slab *

$$p=\frac{E\alpha}{1-\sigma}[Mz+N-\theta]. \quad . \quad . \quad . \quad (1.21)$$

We obtain for the displacements from (1.20),

$$u=Axz+Bx-a''z+c''y+a',$$

and from (1.12), (1.18), and (1.17),

$$v=Ayz+By-b''z-c''x+b'.$$

Also from (1.9) and (1.21),

$$w=\frac{1+\sigma}{1-\sigma}\cdot\alpha\int\theta\,dz-\frac{2\sigma}{1-\sigma}\left(\frac{Az^2}{2}+Bz\right)+w_0,$$

* For a similar equation, based on different assumptions, see "Temperaturverlauf und Wärmespannungen in Verbrennungsmotoren," by G. Eichelberg, *Forschungsarbeiten auf dem Gebiete des Ingenieurwesens*, No. 263 (Berlin, 1923), p. 31.

or

$$w=\frac{1+\sigma}{1-\sigma}.\alpha\int\theta\,dz-\frac{2\sigma}{1-\sigma}\left(\frac{Az^2}{2}+Bz\right)$$
$$-\frac{Ax^2}{2}-\frac{Ay^2}{2}+a''x+b''y+c'$$

using (1.18). Now terms containing a', b', c', and a'', b'', c'', represent a displacement which would be possible in a rigid body *. Since the slab does not rotate we may omit a'', b'', and c''. If the displacements are measured relative to the origin of coordinates (at which u, v, and w are then each zero), we may omit a' and b', while

$$c'+\frac{1+\sigma}{1-\sigma}.\alpha\int\theta\,dz=0,$$

when x, y, and z are zero, *i. e.*,

$$c'+\frac{1+\sigma}{1-\sigma}.\alpha\int\theta\,dz=\frac{1+\sigma}{1-\sigma}.\alpha\int_0^z\theta\,dz.$$

Finally, the displacements are expressed by the equations:

$$\left.\begin{aligned}u&=\alpha x(Mz+N)\\ v&=\alpha y(Mz+N)\\ w&=\alpha\left[\frac{1+\sigma}{1-\sigma}\int_0^z\theta\,dz-\frac{\sigma}{1-\sigma}.z(Mz+2N)-\frac{M}{2}(x^2+y^2)\right]\end{aligned}\right\}. \quad . \quad . \quad . \quad (1.22)$$

Edges free from external loads or restraints.—Since (1.21) does not contain x and y, p cannot vanish at the edges of the slab except when the stress is everywhere zero. Thus (1.21) represents the stresses due to (*a*) variation of temperature across the slab, and (*b*) externally applied stresses, normal to the edges, as given by (1.21). Now according to Saint-Venant's "principle of the elastic equivalence of statically equipollent systems of load" †, the internal stresses due to (*b*), at distances from the edges which are large as compared with the thickness of the slab, depend only on the statical resultant of (*b*), and are independent of its exact mode of application. Hence, if the arbitrary constants, M and N, are found so that the stress distribution (1.21) is statically equivalent to zero force and zero couple, the effect of (*b*) in producing stress will be purely local, and (*b*) may therefore

* Love, *loc. cit.* p. 50.
† Love, *loc. cit.* p. 131.

be removed without affecting the stresses at points well removed from the edges of the slab *. It follows that (1.21) and (1.22) will then give the correct temperature stresses and displacements, except near the edges, in a slab entirely free at the boundaries from external loads or restraints. Thus, for M and N, we have

$$\int_0^c p.dz=0 \quad \text{and} \quad \int_0^c p.z.dz=0.$$

Substituting for p from (1.21), and integrating,

$$\frac{Mc^2}{2}+Nc-\int_0^c \theta\, dz=0,$$

$$\frac{Mc^3}{3}+\frac{Nc^2}{2}-\int_0^c \theta.z\, dz=0,$$

and solving for M and N,

$$\left.\begin{aligned} M&=\frac{12}{c^3}\int_0^c \theta.z\,dz-\frac{6}{c^2}\int_0^c \theta\, dz,\\ N&=\frac{1}{c}\int_0^c \theta\, dz-\frac{Mc}{2}. \end{aligned}\right\} \quad . \quad . \quad . \quad (1.23)$$

Steady Temperature State.—If steady temperature conditions exist in the slab, the temperature gradient is constant, and we may write $\theta=\beta z+\gamma$, in which case (1.21) becomes linear in z. Clearly, when its resultant is zero force and couple, the stress p vanishes at all points. Thus stresses will only occur in a slab with free edges when the temperature gradient is non-uniform.

Other Boundary Conditions.—(*a*) If known external forces are applied to the edges or faces, the stress system due to which can be calculated, the resultant stresses may be found by superimposing this stress system on that given by (1.21) with (1.23). (*b*) If M and N are both zero, the displacements u and v in (1.22) vanish. This is the particular case where movement of the edges in the x and y directions is completely prevented, and the stress equation then becomes

$$p=-\frac{E\alpha\theta}{1-\sigma}. \quad . \quad . \quad . \quad . \quad . \quad (1.24)$$

(*c*) In many practical examples of walls and slabs, movement of the edges is partly restricted, in which case the temperature stresses will be intermediate between the values given by (1.24) and those given by (1.21) with (1.23), for free edges.

* See § 5.

§ 2. *Temperature and Maximum Stress Equations.*

Temperature equation when rate of flow of heat across face $z=0$ *is constant ; face* $z=c$ at zero temperature.

First assume

$$\theta=\left\{1-\frac{z}{c}-\sum_{m=0}^{m=\infty}A_m e^{-\frac{m^2\pi^2h^2t}{4c^2}}.\cos\frac{m\pi z}{2c}\right\}\frac{Qc}{h^2}, \quad (2.1)$$

in which h^2 is the diffusivity of the material of the slab, and Q is a constant. Differentiating (2.1) with respect to z and putting $z=0$, we have

$$\left(\frac{\partial\theta}{\partial z}\right)_{z=0}=-\frac{Q}{h^2},$$

and the rate of flow of heat, q, across unit area of the face $z=0$, is

$$-K\frac{\partial\theta}{\partial z}=\frac{KQ}{h^2}=\rho CQ=\text{a constant},$$

since $h^2=\frac{K}{\rho C}$*, K being the thermal conductivity, ρ the density, and C the specific heat of the slab material. Thus (2.1) satisfies the assumed condition, viz., that q is constant.

By differentiating (2.1) it can easily be seen that the fundamental conduction equation $\frac{\partial\theta}{\partial t}=h^2\frac{\partial^2\theta}{\partial z^2}$ is satisfied. Also when $t=\infty$, $\theta\propto 1-\frac{z}{c}$, and steady conditions obtain.

Again, since we suppose that the slab is at a uniform temperature (taken as zero) at the instant $t=0$ when the heat is initially applied, θ must be zero for all values of z when $t=0$. Thus, from (2.1),

$$0=\left\{1-\frac{z}{c}-\sum_{m=0}^{m=\infty}A_m\cos\frac{m\pi z}{2c}\right\}\frac{Qc}{h^2}, \quad . \quad (2.2)$$

and the coefficient A_m in the Fourier's series is given by

$$A_m=\frac{1}{c}\int_0^{2c}\left(1-\frac{\lambda}{c}\right)\cos\frac{m\pi\lambda}{2c}.d\lambda$$

$$=\frac{4}{m^2\pi^2}\{1-(-1)^m\}=0 \text{ if } m \text{ is even, } =\frac{8}{m^2\pi^2} \text{ if } m \text{ is odd.}$$

Writing $m=2n+1$,

$$A_m=A_{2n+1}=\frac{8}{(2n+1)^2\pi^2},$$

where $n=0, 1, 2, 3 \ldots\ldots \infty$.

* Ingersoll and Zobel, 'Mathematical Theory of Heat Conduction,' p. 3.

Hence (2.1) becomes

$$\theta = \left\{ 1 - \frac{z}{c} - \frac{8}{\pi^2} \sum_{n=0}^{n=\infty} (2n+1)^{-2} e^{\frac{-(2n+1)^2 \pi^2 h^2 t}{4c^2}} \cdot \cos \frac{(2n+1)\pi z}{2c} \right\} \frac{Qc}{h^2} . \quad . \quad . \quad (2.3)$$

Finally, (2.3) clearly satisfies the assumed boundary condition that, when $z=c$, $\theta=0$ for all values of t, and is therefore the required temperature equation.

Constants in Stress Equation.—We now evaluate the integrals in (1.23), substituting for θ from (2.3). We have

$$\left.\begin{aligned} \int_0^c \theta z\, dz &= \frac{c^3 Q}{h^2} \left[\frac{1}{6} - \frac{16}{\pi^3} \sum_0^\infty (2n+1)^{-3} \cdot e^{-(2n+1)^2 \cdot \frac{\pi^2 h^2 t}{4c^2}} \left\{ (-1)^n - \frac{2}{(2n+1)\pi} \right\} \right] \\ \int_0^c \theta\, dz &= \frac{c^2 Q}{h^2} \left[\frac{1}{2} - \frac{16}{\pi^3} \sum_0^\infty (2n+1)^{-3} \cdot e^{-(2n+1)^2 \cdot \frac{\pi^2 h^2 t}{4c^2}} \cdot (-1)^n \right] \end{aligned}\right\} . \quad (2.4)$$

Equations (2.4) enable M and N in (1.23) to be calculated for any given instant, and hence the stress p, at any given distance z from the heated face, may be found from (1.21) and (2.3).

Maximum Compressive Stress.—Of special interest are the maximum stresses occurring in the slab. Inspection of the curves*, showing the stress p at various times and distances from the heated face, shows that the maximum *compressive* stress occurs at the heated face. Putting $z=0$ in (1.21), and substituting for N from (1.23),

$$p_{z=0} = \frac{E\alpha}{1-\sigma} \cdot \left[\frac{4}{c} \int_0^c \theta\, dz - \frac{6}{c^2} \int_0^c \theta z\, dz - \theta \right] .$$

Using (2.4) we find, after some simplification,

$$\left.\begin{aligned} p_{z=0} &= \frac{E\alpha}{1-\sigma} \cdot \frac{8Qc}{\pi^2 h^2} \sum_{n=0}^{n=\infty} B_n \cdot e^{-(2n+1)^2 \cdot \gamma}, \\ \text{in which} \quad B_n &= (2n+1)^{-4} \left\{ (2n+1)^2 + \frac{4}{\pi} (2n+1)(-1)^n - \frac{24}{\pi^2} \right\}, \\ \text{and} \quad \gamma &= \frac{\pi^2 h^2 t}{4c^2} . \end{aligned}\right\}$$

* See fig. 4. . . . (2.5)

The first four values of B_n are : $-0{\cdot}15847$, $0{\cdot}03393$, $0{\cdot}04630$, and $0{\cdot}01567$.

The value of γ which renders the absolute magnitude of the series in (2.5) a maximum is easily found to be $\gamma=0{\cdot}1305$. The series converges rapidly for this value, and it is only necessary to work out the first five terms for a few trial values of γ to obtain this result. Substituting this value in (2.5), we find

$$\text{Maximum of } |p_{z=0}| = \frac{E\alpha}{1-\sigma} \cdot \frac{Qc}{h^2} \cdot 0{\cdot}10278. \quad . \quad (2.6)$$

For the time when this stress occurs we have

$$t = 0{\cdot}0530 \cdot \frac{c^2}{h^2}, \quad . \quad . \quad . \quad . \quad . \quad . \quad (2.7)$$

and for the temperature of the heated face at this time (2.3) gives

$$\theta = 0{\cdot}2596 \cdot \frac{Qc}{h^2}. \quad . \quad . \quad . \quad . \quad . \quad . \quad (2.8)$$

Simplified Equations for Numerical Calculation.—When the slab thickness is infinite the temperature equation takes a simple form which may be used in many cases instead of (2.3) to give the temperature in a slab of finite thickness, without introducing any appreciable error. In such cases the integrals in (1.23) may be expressed very simply, which much facilitates numerical computation of the stresses.

Thus, from the equation for an instantaneous plane source in an infinite solid, we have for a permanent source Q, on the plane $z=0$, bounding a semi-infinite solid *,

$$\theta = \frac{Q}{h\sqrt{\pi}} \int_0^t \frac{1}{\sqrt{t-\tau}} \cdot e^{-\frac{z^2}{4h^2(t-\tau)}} \cdot d\tau.$$

Putting $\zeta = \dfrac{z}{2h\sqrt{t-\tau}}$, we obtain

$$\theta = \frac{Qz}{h^2\sqrt{\pi}} \int_{\frac{z}{2h\sqrt{t}}}^{\infty} \frac{1}{\zeta^2} \cdot e^{-\zeta^2} d\zeta,$$

from which, integrating by parts,

$$\theta = \frac{Qz}{h^2} \left[\frac{e^{-\beta^2}}{\beta\sqrt{\pi}} - \{1-\phi(\beta)\} \right], \quad . \quad . \quad . \quad (2.9)$$

* Ingersoll and Zobel, *loc. cit.* p. 98.

where

$$\beta=\frac{z}{2h\sqrt{t}}, \quad \phi(\beta)=\frac{2}{\sqrt{\pi}}\int_0^{\beta} e^{-\beta^2}d\beta,$$

$\phi(\beta)$ being the probability integral.

(2.9) may be written

$$\theta=2Q\sqrt{\frac{t}{h^2}}.\psi(\beta), \quad . \quad . \quad . \quad . \quad (2.10)$$

from which θ may easily be calculated with the aid of the following table:

β.	$\psi(\beta)$.	β.	$\psi(\beta)$.	β.	$\psi(\beta)$.
0·0	0·5642	0·6	0·1559	1·4	0·0127
0·1	·4698	0·7	·1201	1·6	·0058
0·2	·3866	0·8	·0912	1·8	·0025
0·3	·3142	0·9	·0682	2·0	·0010
0·4	·2521	1·0	·0503	2·2	·0004
0·5	·1996	1·2	·0260	2·4	·0001

(2.10) will give the correct temperatures in a slab of finite thickness c, if on putting $\beta=\frac{c}{2h\sqrt{t}}$, θ is very nearly zero, thus satisfying the assumed boundary conditions.

To find the integrals in (1.23), we note that in this case

$$\int_0^c \theta\,dz=\int_0^{\infty}\theta\,dz, \text{ very nearly, } =4Qt\int_0^{\infty}\psi(\beta)d\beta$$

and

$$\int_0^c \theta z\,dz=\int_0^{\infty}\theta z\,dz, \text{ very nearly, } =8Qht^{3/2}\int_0^{\infty}\beta\psi(\beta)d\beta$$

from (2.10). By repeated integration by parts, it can be shown that

$$\int_0^{\infty}\psi(\beta)d\beta=\frac{1}{4}, \quad \int_0^{\infty}\beta\psi(\beta)d\beta=\frac{1}{6\sqrt{\pi}}.$$

Hence, finally,

$$\int_0^c \theta\,dz=Qt, \quad \int_0^c \theta z\,dz=\frac{4}{3\sqrt{\pi}}Qht^{3/2}, \quad . \quad (2.11)$$

from which M and N may easily be obtained.

§ 3. *Estimation of the Constant* Q *in the Temperature Equation, to represent the Heating Effect of Solar Radiation.*

We proceed to show that (2.3) will approximately represent the heating of a flat horizontal slab, exposed continuously to the sun's rays.

The effect of the variation in the sun's position may be allowed for as follows:—We have, from the astronomical spherical triangle *,

$$\cos\zeta = \sin\phi \,.\, \sin\delta \,.\, + \cos\phi \,.\, \cos\delta \,.\, \cos \mathrm{P}, \qquad (3.1)$$

where ζ is the zenith distance, δ the declination, and P the hour-angle of the sun's centre, and ϕ is the latitude of the place where the slab is situated.

The proportion of solar radiation transmitted by the atmosphere, on a perfectly clear day, varies considerably with the distance of the sun from the zenith, and may be estimated with sufficient accuracy for our present purpose by the formula †

$$\mathrm{S} = \mathrm{S}' a^{\sec.\zeta},$$

in which S′ is the solar constant, S is the fraction of solar radiation transmitted by the atmosphere, and a is the transmission coefficient for a vertical sun.

If r is the proportion of radiant heat incident to the slab which is reflected from its surface, the quantity, q_2, of heat absorbed by the slab is

$$\frac{\mathrm{S}}{60}(1-r)\cos\zeta \text{ per sec. per sq. cm.,}$$

or

$$q_2 = \frac{\mathrm{S}'}{60} a^{\sec.\zeta} \,.\, (1-r)\cos\zeta. \quad . \quad . \quad . \quad . \quad (3.2)$$

Now assuming that the loss of heat, by radiation and convection, obeys Newton's law of cooling, heat radiated from the exposed face $z=0$, at temperature $\theta_0 = \lambda\theta_0$ cal. per sec. per sq. cm., λ being the emissivity constant.

If q is the resultant inward flow of heat across the face $z=0$, then

$$q = q_2 - \lambda\theta_0.$$

* 'Topographical Surveying,' by Close and Winterbotham, p. 170 (1925).

† Preston, 'Theory of Heat,' 4th ed. p. 584. The formula only strictly applies to monochromatic radiation.

Now, assuming tentatively that q is approximately constant, we have from (2.10), putting

$$z = \beta = 0, \quad \text{and} \quad \psi(0) = \frac{1}{\sqrt{\pi}},$$

$$\theta_0 = \frac{2Q}{h\sqrt{\pi}} \cdot \sqrt{t} = \frac{2q}{\rho C h \sqrt{\pi}} \cdot \sqrt{t}.$$

Then, finally,

$$q = \frac{q_2}{1 + \eta\sqrt{t}}, \text{ where } \eta = \frac{2\lambda}{\rho C h \sqrt{\pi}}. \quad . \quad . \quad (3.3)$$

If we now calculate values of q_2 from (3.1) and (3.2), for assumed hour-angles of the sun, and substitute these in turn in (3.3), we find that by choosing a suitable value of P for which the time t is 0, q is approximately constant with respect to P, as assumed above.

For example, assume

$S' = 1{\cdot}93$ cal. per sq. cm. per min.*

$a = 0{\cdot}75$ for a very clear sky.

$\phi = 51\frac{1}{2}°$ at London.

$\delta = +23\frac{1}{2}°$ at midsummer.

$r = 0{\cdot}17$ for granolith (pavement)†.

$\lambda = 0{\cdot}00025$ cal. per sq. cm. per sec. per degree C.*

$h^2 = 0{\cdot}0056$, $\rho = 2{\cdot}3$, $C = 0{\cdot}21$, C.G.S. units, for concrete.

From (3.1) and (3.2), we obtain the values from which curve 1 of q_2 and P is drawn in fig. 2. Taking $t = 0$ when $P = 5$ hours (before noon), and $q = 0{\cdot}00783$ cal. per sq. cm. per sec., we obtain from (3.3) the values of q_2 and t from which curve 2 is drawn. The agreement between the two curves, though not very close, is satisfactory when the uncertainty as to the actual atmospheric conditions and their effect upon the heat absorbed by the slab is remembered.

It has been assumed in fig. 2 that no heat is received by the slab before $P = 5$ hours. In any event the heat from the rising sun is small, and to comply strictly with this assumption, we may suppose the sun to be obscured by cloud before this time.

* Smithsonian Physical Tables, pp. 181, 252.

† Coblentz, in U.S. Bureau of Standards, ix. p. 283 (1912).

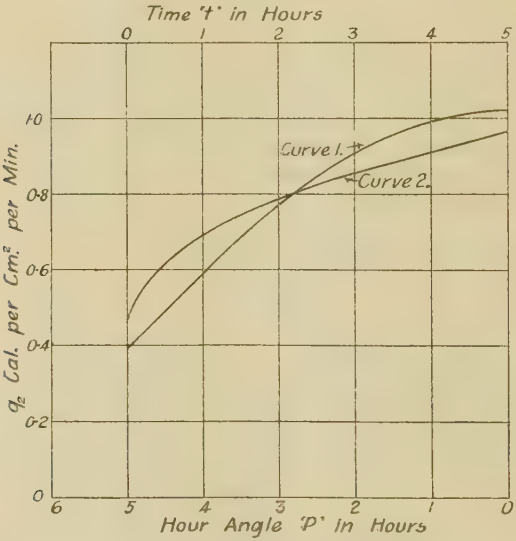
For the constant Q in (2.3), we have

$$Q = \frac{q}{\rho C} = \frac{\cdot 00783}{\cdot 21 \times 2.3} = 0\cdot 0162^\circ \text{ C. cm. sec.}^{-1},$$

the value used in § 4, which follows.

The heat received by a *vertical* wall could be estimated as above, only using an equation corresponding to (3.1) to give the angle between the normal to the wall and the sun's rays at any instant. We should find that q in this case would be

Fig. 2.



considerably smaller than the above value for a horizontal slab.

§ 4. *Numerical Values of the Temperatures and Stresses.*

1. *Horizontal Concrete Slab, 18 inches thick, unrestrained at the edges.*

The following values of the constants are assumed:

$\alpha = 10^{-5}$ per degree C.*

$\sigma = 0\cdot 15$ †.

* 'American Civil Engineers' Handbook,' p. 565 (1930).

† *Loc. cit.* p. 562.

$E = 3 \times 10^6$ lb. per sq. in. (for 3000 lb. concrete)*.

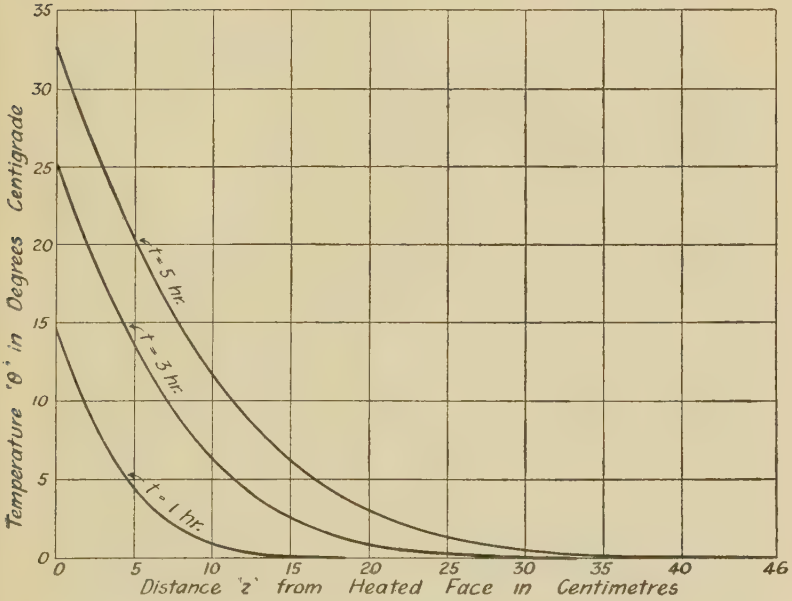
$Q = 0{\cdot}0162°$ C. cm. sec.$^{-1}$.

$h^2 = 0{\cdot}0057$ cm.^{2} sec.$^{-1}$†. $c = 46$ cm.

Using these values, (2.10) becomes

$$\theta = 0{\cdot}429\sqrt{t}.\psi(\beta), \text{ in which } \beta = \frac{z}{0{\cdot}151\sqrt{t}},$$

Fig. 3.



Curves showing the temperature distribution through a horizontal concrete slab, after exposure to the sun for various times.

t being in seconds and z in centimetres. The curves in fig. 3, showing the temperature distribution over the thickness of the slab, were calculated from this equation and the table in § 2, for times of 1, 3, and 5 hours from the commencement of heating. The rise in temperature of the face $z=0$ is only $0{\cdot}02°$ C., when $t=5$ hours. Thus the use of the simplified equations (2.10) and (2.11), as explained in § 2,

* Faber, Proc. Inst. C. E. ccxxv. p. 59.

† Ingersoll and Zobel, *loc. cit.* p. 162.

is justified in this case. The values of the constants in the stress equation are:

t.	$-M$. (°C./cm.).	N. (°C.).
h. m.		
5 29	0·593	20·58
5 00	0·554	19·08
3 00	0·3686	12·29
1 00	0·1412	4·52

For each of the times in the table, the stresses for 18 values of z, ranging from 0 to 46 cm., were calculated, and from these the stress curves in fig. 4 were drawn. The curve for $t=5^h\ 29^m$ is not shown, since it could not be clearly distinguished from the 5 hour curve.

Maximum Compressive Stress.—It appears from the curves that the maximum compressive stress occurs at the heated face, as assumed in deriving (2.6). With the assumed numerical values, we have from (2.6): Maximum compressive stress = 33·2 kg. per sq. cm. (472 lb. per sq. in.), and from (2.7), the time at which this stress occurs is 19,700 seconds, or $5^h\ 28^m\ 20^s$. As a check on this stress we may put $z=0$ in (1.21) and N = 20·58, corresponding to $t=5^h\ 29^m$ in the above table based on the simplified equations, and we find that the stress at the heated face is 33·0 kg. per sq. cm. The maximum compressive stress is therefore about one-sixth of the ultimate strength of the concrete.

Maximum Tensile Stress.—The curves in fig. 4 also show that the maximum tensile stress, for the times considered, is 12·4 kg. per sq. cm. (177 lb. per sq. in.), and this occurs at a distance from the heated face equal to 0·4 times the thickness of the slab. The maximum tensile stress at the instant when the maximum compressive stress occurs—*i.e.*, 29 min. later than the time for which the curve is drawn—is 12·8 kg. per sq. cm., the temperature at $z=19·1$ cm., where this occurs, being 4·1° C. Though the true maximum probably occurs at a slightly later time, it seems clear without further calculation that it will not appreciably exceed those values just given. Moreover, the intensity of solar heat assumed is unlikely to endure for much longer than 5 hours.

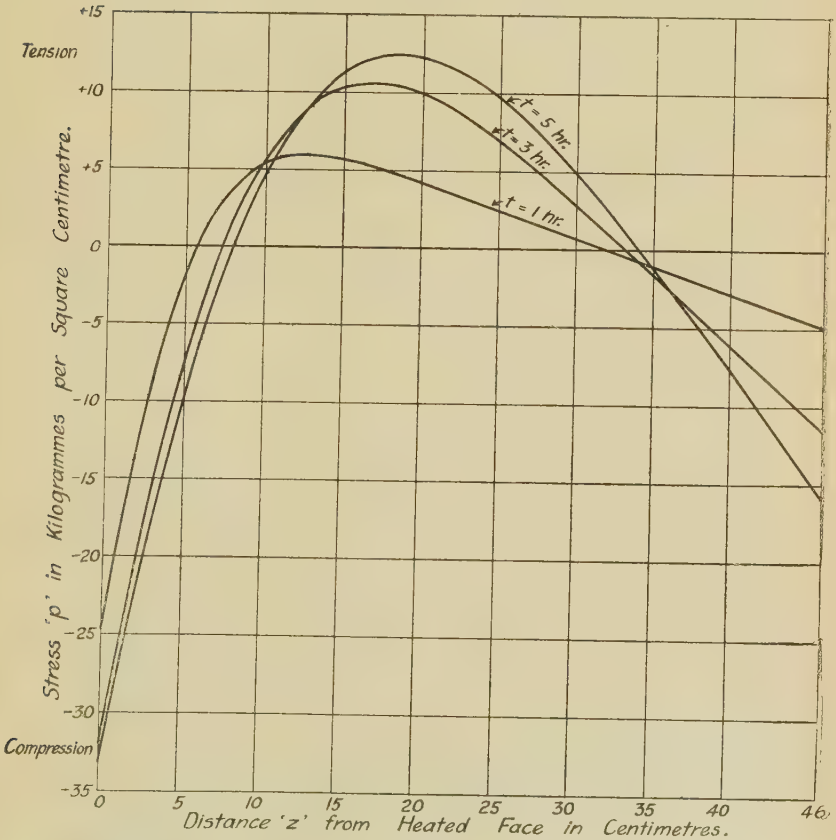
The maximum tensile stress is thus about half the ultimate tensile strength of the concrete.

Variation of Maximum Compressive Stress with c.—From (2.6) it appears that the maximum compressive stress is

proportional to the thickness c of the slab, provided that the time (from (2.7)) when this occurs does not exceed the maximum duration of the source of heat. In the present example, if maximum duration = 6 hours = time for maximum stress,

$$c = 48{\cdot}2 \text{ cm. (from (2.7))}$$
$$= 19 \text{ in.}$$

Fig. 4.



Curves showing the distribution of stress through a horizontal concrete slab in which the temperature variation is as shown in fig. 3.

and the above relation applies for slabs 19 in. or less in thickness.

Restrained Edges.—If the edges are so restrained that expansion in the x and y directions is completely prevented, the stresses are proportional to the ordinates of the temperature curves in fig. 3, and in this case are practically independent of the slab thickness (except for thin slabs). From (1.24) the maximum stress for $t=5$ hours is 80·5 kg. per sq. cm. (1146 lb. per sq. in.).

2. *Horizontal Slab of Common Brickwork.*—As a second example, consider the maximum temperature stresses in a horizontal slab of brick-masonry of the same thickness as the concrete slab in the previous example. The approximate values of the constants are in this case: $\alpha=5{\cdot}4\times10^{-6}$*, $\sigma=0{\cdot}25$†, $E=1{\cdot}4\times10^{5}$‡, $Q=0{\cdot}0162$, $h^2=0{\cdot}005$§, $c=46$ (c.g.s. units).

The equations give the following values:

Maximum compressive stress = 15·5 kg. per sq. cm.
= 0·18 Ultimate Strength,

taking the latter as 84·4 kg. per sq. cm.

Corresponding time = $6^{h}\ 13^{m}\ 20^{s}$, and
temperature of heated face = 38·6° C.

Maximum tensile stress = 5·91 kg. per sq. cm.,
(at the same time)
for which $z=19{\cdot}1$ cm., $\theta=4{\cdot}7°$ C.

Conclusions from Examples.—The maximum compressive temperature stress, occurring in a concrete or brickwork slab, exposed horizontally to the sun's rays, may be of the same order as the permissible working stress. Since this stress occurs at the outer face, and is of the same kind as the stresses due to dead weight and loading, it would seem to be a possible factor causing disintegration of the slab surface.

The maximum tensile stress, occurring near the middle of the slab, and being for thick slabs of the same order as the ultimate tensile strength of the material, may possibly cause cracks to form, perpendicular to the faces, in the interior of the slab. Though these cracks would cause some redistribution of stress, it seems unlikely that they can be of serious consequence. Moreover, loading and dead-weight stresses usually tend to neutralize the tension due to temperature changes, in which case the resultant tension is small.

* 'American Handbook,' p. 891.
† 'Dictionary of Applied Physics,' i. p. 176.
‡ 'American Handbook,' p. 891.
§ Ingersoll and Zobel, *loc. cit.* p. 162.

§ 5. *Limitations of Theory near the Edges of the Slab.*

In obtaining the stresses in a slab with free edges, the local effect of the external boundary stresses (having zero moment and zero resultant), which are required to make the solution strictly accurate at all points, has been neglected. We may obtain some idea of the limits beyond which the effect of these external stresses is negligible by referring to the stress distribution in a tension member of finite width and infinite length pierced by a circular hole. An exact solution of this problem, when the hole is central, has been obtained by R. C. J. Howland *. Referring to the stress diagrams on page 85 of Dr. Howland's paper, for the case when the diameter of the hole is equal to half the width of the plate, it appears that the local disturbance due to the hole becomes negligible at a distance from the centre equal to one and a half times the width of the plate, *i. e.*, the tension across the plate is practically uniform at this distance from the hole. Now comparing this case with the stress distribution in fig. 4 of the present paper, and supposing that a cross-section through the centre of the hole corresponds with the edge of the slab in our problem, we see that this result may be taken as indicating the rate at which the boundary compression, applied to the outer parts of the edge will be distributed.

Reference may now be made to the results of some optical stress measurements in flat tension members with enlarged ends, an account of which is given in ' Photo Elasticity ' by Coker and Filon. It there appears from fig. 7·027 on page 566, that when a plate abruptly enlarges to twice its width, the stress becomes uniformly distributed over the enlarged cross-section at a distance of about three-quarters of the enlarged width from the change of section. We may therefore infer that the tension over the central part of the edge (for which $7 < z < 33$, in fig. 4) will become similarly distributed.

We conclude from these considerations that the externally applied tension and compression become distributed—and since they are equal, therefore neutralize each other—within a distance of about one and a half times the thickness of the slab. Hence the solution in § 1 will accurately give the temperature stresses and displacements, in a slab free from external stress or restraint, at points further from the edges than this distance.

* Phil. Trans. A, ccix. pp. 49–86 (1930).

X. *Vibrations of Turbine Blades with Shrouding.*

By KATSUTADA SEZAWA, *Tokyo Imperial University* *.

THE problem of the natural vibrations of an isolated turbine blade in rotation has been studied by many mathematicians and engineers in different manners. Among these the result due to Ono † concerning the case of a blade of variable cross-section vibrating mainly in its flexible direction seems to be remarkable. More general cases of vibrations of a bar of variable cross-section but not in rotation have been studied by Nicholson ‡, Wrinch §, Ono ‖, and others. The problem of the vibrations of a group of turbine blades held together by shrouding, however, does not appear to have received much attention. Although Schwerin ¶ has written on the vibrations of a group of blades, yet, owing to certain apparent particularities on his part, his result is open to some doubts. I have now renewed this problem considered from a more theoretical standpoint and arrived at some results which may be reliable, at least in the case where our idealized conditions are fulfilled. The problem is, of course, connected with the case of an axial-flow turbine.

Fundamental assumptions involved in the present paper are as follows :—In the first place, the sections of turbine blades and shrouding are uniform throughout their respective lengths. It is, of course, possible to deal with the problem of blades with variable cross-section : yet I have taken up the simplest case as mentioned above in order to avoid the complexity of the problem. Secondly,

* Communicated by the Author.

† A. Ono, "Vibrations of Turbine Blades," Journ. Soc. Mech. Eng. Tokyo, xxvii. pp. 467–479 (1924); Geiger u. Scheel, *Handb. d. Phys.* iv. p. 370 (Berlin, 1928); Timoshenko, 'Vibration Problems in Engineering,' p. 267 (New York, 1928).

‡ J. W. Nicholson, "The Lateral Vibrations of Bars of Variable Section," Proc. Roy. Soc. Lond. xciii. pp. 506–519 (1917); "The Lateral Vibrations of Sharply-pointed Bars," Proc. Roy. Soc. Lond. xcvii. pp. 172–181 (1920).

§ Dorothy M. Wrinch, "On the Lateral Vibrations of Rods of Variable Cross-section," Phil. Mag. xlvi. pp. 273–291 (1923).

‖ A. Ono, "Lateral Vibrations of Tapered Bars," Journ. Soc. Mech. Eng. Tokyo, xxviii. pp. 429–441 (1925).

¶ E. Schwerin "Ueber die Eigenfrequenzen der Schaufelgruppen von Dampfturbinen," *Verh. 2-int. Kongr. tech. Mech.* pp. 207–212 (1926); *Z. f. tech. Phys.* viii. pp. 312–319 (1927).

the series of blades in a group is assumed infinite, again for the sake of simplicity. As the number of blades bound in a group by means of shrouding is large in usual cases, this assumption is not violent. In the third place, the deflexion vibrations of blades in one group are mainly in the direction of the least moment of inertia of the section of each blade. Hence, the effective consecutive length of the shrouding, which is intercepted between two neighbouring blades and in the most flexible direction of these blades, is smaller than that length in the peripheral direction of the turbine wheel. It is to be remarked that the small vibrations of the group in the direction transverse to that of the least moment of inertia of the blade section is not difficult from the nature of the problem and that the problem is not much important in the case of blades with shrouding, so that I have not treated of such a special case.

It is, in fact, desirable to solve the problem of the vibrations that involve the flexural rigidity of the group as well as the centrifugal effect. But, on account of the difficulty of the solution of the differential equation of the vibration, I have followed the method of the combination of frequency equations of a rotating elastic body which has already been employed by Southwell *, Ono †, Suyehiro ‡, and many others; that is, if $p'/2\pi$ is the frequency of an elastic body without rotation, and $p''/2\pi$ that for the same body in rotation when the effect of the flexural rigidity is neglected, the resulting frequency $p/2\pi$ of the body, when the flexural rigidity and the effect of rotation are both taken into consideration, is expressed by

$$p^2 = p'^2 + p''^2.$$

I. *Effect of Rotation being neglected.*

In this paper the following notations are employed:

ρ_1, ρ_2 = densities of blade and shrouding respectively;

E_1, E_2 = Young's moduli of blade and shrouding respectively;

* R. V. Southwell, "On the Free Transverse Vibrations of a Uniform Circular Disk clamped at its Centre, and on the Effect of Rotation," Proc. Roy. Soc. Lond. ci. pp. 133–153 (1922); also his other papers.

† A. Ono, *loc. cit.*

‡ K. Suyehiro, "On the Gyroscopic Vibration of Marine Turbine Disks," Journ. Soc. Nav. Arch. Tokyo, xxxix. pp. 121–128 (1926).

a_1, a_2=effective areas of blade and shrouding respectively;

k_1, k_2=radii of gyration of the section of blade and shrouding respectively;

l_1, l_2=effective lengths of blade and of the intercepted portion of shrouding respectively;

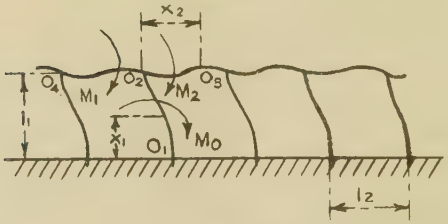
y_1, y_2=transverse deflexions of blade and shrouding at x_1 and x_2;

M_0, M_1, M_2=bending moments of blade and shrouding at a joint O_2;

R=radius of turbine wheel;

ω=angular velocity of turbine wheel.

Fig. 1.



Diagrammatic sketch of blades with shrouding.

Let us first consider the case $\omega=0$, the effect of curvature being neglected. The equation of motion of the portion O_1O_2 is expressed by

$$E_1{k_1}^2\frac{\partial^4 y_1}{\partial {x_1}^4}+\rho_1\frac{\partial^2 y_1}{\partial t^2}=0. \quad . \quad . \quad . \quad . \quad (1)$$

If we write $y_1=u_1\cos p't$, then

$$\frac{d^4u_1}{d{x_1}^4}={m_1}^4u_1, \quad . \quad . \quad . \quad . \quad . \quad . \quad (2)$$

where

$${m_1}^4=\frac{\rho_1 p'^2}{E_1{k_1}^2}. \quad . \quad . \quad . \quad . \quad . \quad . \quad (3)$$

The solution of (2) is

$$u_1=A_1\cos m_1x_1+B_1\sin m_1x_1+C_1\cosh m_1x_1+D_1\sinh m_1x_1. \quad . \quad . \quad (4)$$

The boundary conditions at O_1, which is the end of the blade clamped to the disk, are written

$$x_1=0\,;\quad u_1=0\,;\quad \frac{du_1}{dx_1}=0. \quad . \quad . \quad . \quad . \quad . \quad (5)$$

Substituting from (4) in (5), we get

$$u_1=A_1(\cos m_1x_1-\cosh m_1x_1)+B_1(\sin m_1x_1-\sinh m_1x_1). \quad (6)$$

The boundary conditions at the other end, O_2, are

$$x_1=l_1\,;\quad -E_1a_1k_1^2\frac{d^2u_1}{dx_1^2}=M_0\,;$$

$$-E_1a_1k_1^2\frac{d^3u_1}{dx_1^3}=\rho_2a_2l_2p'^2u_1. \quad (7),\ (8)$$

The equation of the transverse motion of the member O_2O_3 (or O_4O_2) is expressed by

$$E_2k_2^2\frac{\partial^4y_2}{\partial x_2^4}+\rho_2\frac{\partial^2y_2}{\partial t^2}=0. \quad . \quad . \quad . \quad . \quad (9)$$

If we write $y_2=u_2\cos p't$, then

$$\frac{d^4u_2}{dx_2^4}=m_2^4u_2, \quad . \quad . \quad . \quad . \quad . \quad . \quad (10)$$

where

$$m_2^4=\frac{\rho_2p'^2}{E_2k_2^2}. \quad . \quad . \quad . \quad . \quad . \quad . \quad (11)$$

The solution of (10) is

$$u_2=A_2\cos m_2x_2+B_2\sin m_2x_2+C_2\cosh m_2x_2+D_2\sinh m_2x_2.$$
$$. \quad . \quad . \quad (12)$$

Assume that the connexion between blade and shrouding is perfectly rigid; then the conditions at the joint $x_2=0$, $x_1=l_1$ are

$$u_2=0,\quad \frac{du_2}{dx_2}=\frac{du_1}{dx_1},\quad -E_2a_2k_2^2\frac{d^2u_2}{dx_2^2}=M_2\,; \quad (13),\ (14),\ (15)$$

and the conditions at the joint $x_2=l_2$, $x_1=l_1$ are

$$u_2=0,\quad \frac{du_2}{dx_2}=\frac{du_1}{dx_1},\quad -E_2a_2k_2^2\frac{d^2u_2}{dx_2^2}=M_1. \quad (16),\ (17),\ (18)$$

The condition of the fixing moment at O_2 (or O_3) is written

$$M_0+M_1-M_2=0. \quad . \quad . \quad . \quad . \quad . \quad (19)$$

Applying the conditions and the relations in (7), (8), (13), (14), (15), (16), (17), (18), (19) to the equations (6) and (12), it is possible to find the ratios of the constants A_1, B_1, A_2, B_2, C_2, D_2, besides M_0, M_1, M_2, and to determine the values of m_1 and m_2 which will give us the frequency $p'/2\pi$.

Write

$$m_1 l_1 \equiv \alpha, \quad m_2 l_2 \equiv \beta, \quad \frac{\rho_2 a_2 l_2 p'^2}{E_1 a_1 k_1^2 m_1^3} \equiv \gamma, \quad \frac{m_2}{m_1} \equiv \xi,$$

$$\frac{E_2 a_2 k_2^2 m_2^2}{E_1 a_1 k_1^2 m_1^2} \equiv \eta ; \quad (20)$$

then, from the conditions (8), we get

$$A_1[(\sin\alpha - \sinh\alpha) + \gamma(\cos\alpha - \cosh\alpha)] = B_1[(\cos\alpha + \cosh\alpha) - \gamma(\sin\alpha - \sinh\alpha)]. \quad (21)$$

From (13) and (14) we find

$$(B_2 + D_2)\xi = -A_1(\sin\alpha + \sinh\alpha) + B_1(\cos\alpha - \cosh\alpha), \quad (22)$$

while (16) gives us

$$A_2(\cos\beta - \cosh\beta) + B_2 \sin\beta + D_2 \sinh\beta = 0. \quad (23)$$

By means of (17) and (22) we obtain

$$A_2(\sin\beta + \sinh\beta) + B_2(1 - \cos\beta) + D_2(1 - \cosh\beta) = 0. \quad (24)$$

Finally, using the relations (7), (15), (18) and the condition (19), we get

$$A_1(\cos\alpha + \cosh\alpha) + B_1(\sin\alpha + \sinh\alpha) + A_2\eta(-2 + \cos\beta + \cosh\beta) + B_2\eta \sin\beta - D_2\eta \sinh\beta = 0. \quad (25)$$

Eliminating A_1, B_1, A_2, B_2, D_2 between (21), (22), (23), (24), (25), we arrive at the relation

$$2\{(\cos\alpha \sinh\alpha + \sin\alpha \cosh\alpha) + \gamma(\cos\alpha \cosh\alpha - 1)\} \cdot \{\sin\beta(\cosh\beta - 1) + \sinh\beta(1 - \cos\beta)\} = \frac{\xi}{\eta}\{(\cos\alpha \cosh\alpha + 1) + \gamma(\cos\alpha \sinh\alpha - \sin\alpha \cosh\alpha)\} \cdot (\cos\beta \cosh\beta - 1). \quad (26)$$

From this equation and the relations in (20), we can find the frequency $p'/2\pi$ or the period of vibrations $2\pi/p'$.

When the stiffness of the blade is very high compared with that of the shrouding, the condition $\eta/\xi = 0$ holds, so that

$$\cos\beta \cosh\beta - 1 = 0. \quad . \quad . \quad . \quad . \quad (27)$$

This is the frequency equation of the shrouding whose both ends are clamped in position and inclination. In a special case, where there is no shrouding, we know that $\eta/\xi=0$ and $\gamma=0$, so that we get

$$\cos\alpha\cosh\alpha+1=0. \quad . \quad . \quad . \quad . \quad (28)$$

This is the frequency equation of a clamped-free bar and gives the equation of a blade with a free tip.

When the stiffness of the shrouding is very high in comparison with that of the blade and the mass of the shrouding is large, we may put $\xi/\eta=0$ and $1/\gamma=0$, so that

$$\cos\alpha\cosh\alpha-1=0. \quad . \quad . \quad . \quad . \quad . \quad (29)$$

This is the frequency equation of the blade whose both ends are clamped in position and in slope.

TABLE I.

Values of $\left(\frac{\rho_1 p'^2}{E_1 k_1^2}\right)^{\frac{1}{4}} l_1$ corresponding to different a_1/a_2 and l_1/l_2.

a_1/a_2.	1/10.	1/2.	1.	2.	4.	6.	9.
$l_1/l_2=\begin{cases}10\\5\end{cases}$	1·717 —	2·13 1·98	2·226 2·096	2·235 2·131	2·086 1·965	1·966 1·896	1·879 —

As it is very difficult to solve (26) in general, I have employed the method of trial and error to get frequencies in various cases. Again, I have limited the calculation to the determination of the fundamental free vibrations of the cases $l_1/l_2=10$ and $=5$, besides the conditions that $\rho_1=\rho_2$ and $E_1=E_2$ and the effective breadth of the shrouding perpendicular to its flexible direction is the same as that of the blades. Such restricted examples, nevertheless may include almost all probable cases of actual turbine blades. The result is shown in Table I. and fig. 2.

From the result of the calculation it may be seen that the increase of the natural frequency owing to the existence of the shrouding is a maximum when the ratio of the area of the blade to that of the shrouding is in the neighbourhood of 1 or 2; and that the increase of $p'/2\pi$ in this vicinity is roughly 50 per cent. of that without shrouding. When the ratio of the areas is 6 or 7, the attachment of

the shrouding is of no influence on the period of the vibrations, though it may be of use in the prevention of the mutual impact of blades. When the area of the blade is less than half that of the shrouding, the frequency is extremely lowered, tending to zero value as the area of the shrouding becomes larger and larger.

Fig. 2.

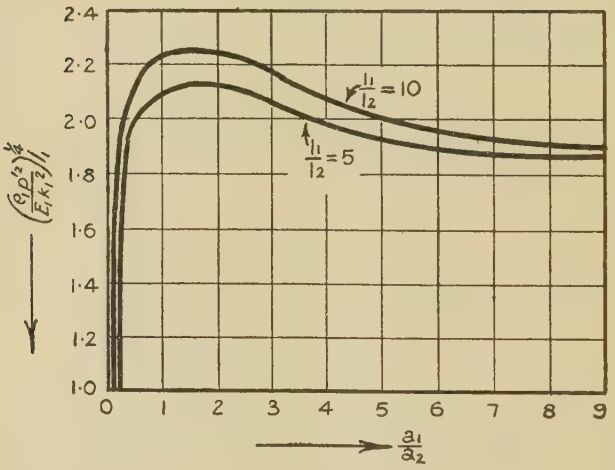


Fig. 3.

II. *Effect of Flexural Rigidity being neglected.*

The problem is the same as that in which the shrouding is concentrated at one end of a flexible clamped-free bar rotating with angular velocity ω (fig. 3).

Take the axis of x along a blade, its origin being at the foot of that blade. The centrifugal force acting at the blade section x is denoted by

$$\rho_2 a_2 l_2 \omega^2(\mathrm{R}+l_1)+\rho_1 a_1 \omega^2 \int_x^{l_1}(\mathrm{R}+x)dx. \quad . \quad . \quad (30)$$

The equation of the transverse vibrations of an elementary portion of the blade is expressed by

$$\rho_1 a_1 \frac{\partial^2 y_1}{\partial t^2} = \frac{\partial}{\partial x}\left[\left\{\rho_2 a_2 l_2 \omega^2 (\mathrm{R}+l_1) + \rho_1 a_1 \omega^2 \int_x^{l_1} (\mathrm{R}+x)dx\right\}\frac{\partial y}{\partial x}\right], \quad . \quad . \quad . \quad . \quad (31)$$

where y is the transverse displacement of the blade at the section x. Writing $y=u \cos p''t$, this equation is transformed to

$$\left[\frac{\rho_2 a_2 l_2}{\rho_1 a_1}(\mathrm{R}+l_1)+\mathrm{R}(l_1-x)+\frac{{l_1}^2-x^2}{2}\right]\frac{d^2u}{dx^2} - (\mathrm{R}+x)\frac{du}{dx}+\frac{p''^2}{\omega^2}u=0. \quad (32)$$

Put $$z=\frac{\mathrm{R}+x}{\mathrm{R}+l_1}, \quad \zeta^2=1+\frac{2\rho_2 a_2 l_2}{\rho_1 a_1(\mathrm{R}+l_1)}, \quad . \quad . \quad . \quad (33)$$

then the equation (32) reduces to

$$\left\{1-\left(\frac{z}{\zeta}\right)^2\right\}\frac{d^2u}{dz^2}-\frac{2z}{\zeta^2}\frac{du}{dz}+\frac{2p''^2}{\zeta^2\omega^2}u=0, \quad . \quad (34)$$

the solution of which is

$$u=\mathrm{AP}_n\left(\frac{z}{\zeta}\right)+\mathrm{BQ}_n\left(\frac{z}{\zeta}\right), \quad . \quad . \quad . \quad (35)$$

provided $$\frac{2p''^2}{\omega^2}=n(n+1). \quad . \quad . \quad . \quad . \quad . \quad (36)$$

The general solution of the vibrations is therefore

$$y=\left\{\mathrm{AP}_n\left(\frac{z}{\zeta}\right)+\mathrm{BQ}_n\left(\frac{z}{\zeta}\right)\right\}\cos p''t. \quad . \quad . \quad (37)$$

This type of solution is quite similar to that which Ono obtained for the problem of a blade of variable cross-section and with no concentrated mass. Thus, in the present treatment I am more or less indebted to his paper.

Now, two kinds of solutions, namely, $\mathrm{P}_n(z/\zeta)$ and $\mathrm{Q}_n(z/\zeta)$, in (37) are necessary for the determination of the problem of a bar with prescribed conditions at both ends. But, as the nature of the bar at its free end is implicitly involved in the equation of (31), and, again, as the numerical treatment of $\mathrm{Q}_n(z/\zeta)$ is not so easy, I have taken the solution of (34) approximately in the form

$$y=\mathrm{AP}_n\left(\frac{z}{\zeta}\right)\cos p''t \quad . \quad . \quad . \quad . \quad . \quad (38)$$

in lieu of (37). The solution contains therefore some constraint, but as the constraint is merely connected with the bending moment of the bar at its free end, the frequency equation is not much affected. In fact, such an approximation is customary in the problems of string and membrane which are more general cases of the example of the present section.

From the present category the boundary condition is merely given by

$$x=0\,;\quad y=0. \qquad . \quad . \quad . \quad . \quad . \quad (39)$$

Substituting from (38) in (39), we get

$$\mathrm{P}_n\Big(\frac{\mathrm{R}}{\mathrm{R}+l_1}\Big/\sqrt{1+\frac{2\rho_2 a_2 l_2}{\rho_1 a_1(\mathrm{R}+l_1)}}\Big)=0, \quad . \quad . \quad (40)$$

where $n(n+1)=2p''^2/\omega^2$.

TABLE II.

Values of l_1/R corresponding to different p''^2/ω^2 and l_2a_2/l_1a_1.

p''^2/ω^2.		3.	6.	10.	15.	21.
$\frac{l_2a_2}{l_1a_1}=$	0	·736	·291	·1615	·104	·073
	1/40	·719	·285	·160	·102	·071
	1/20	·698	·280	·153	·099	·069
	1/10	·678	·263	·149	·097	·068
	1/5	·617	·250	·136	·090	·061
	1	·410	·151	·082	·052	·038

For solving this equation I have again employed the method of trial and error and found the relation between l_1/R, l_2a_2/l_1a_1, and p''^2/ω^2, the result being shown in Table II. and fig. 4.

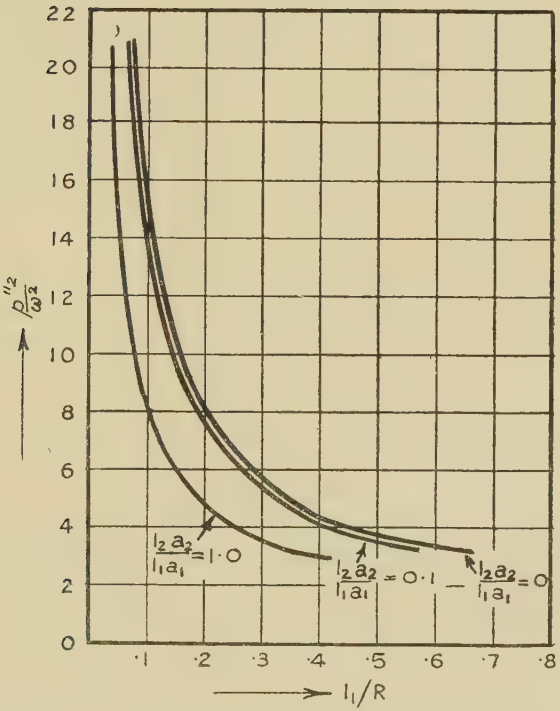
From these results it may be seen that the addition of the shrouding decreases, though very little, the frequency of the natural vibrations of the blade in the case where the effect of the flexural rigidity is neglected. The larger the mass of the shrouding, the more the frequency of the vibrations decreases. It may again be seen that, the greater the radius of the turbine disk, the higher becomes the natural frequency.

III. *Concluding Remark and Example of Application.*

The exact method of dealing with the problem of vibrations of the turbine blades is to solve the equation

$$\rho_1 a_1 \frac{\partial^2 y_1}{\partial t^2} + \mathrm{E}_1 a_1 k_1^2 \frac{\partial^4 y_1}{\partial x^4} - \frac{\partial}{\partial x}\left[\left\{\rho_2 a_2 l_2 \omega^2 (\mathrm{R}+l_1) + \rho_1 a_1 \omega^2 \int_x^{l_1} (\mathrm{R}+x)dx\right\} \frac{\partial y_1}{\partial x}\right] = 0 \quad (41)$$

Fig. 4.



for the blade element, and a somewhat modified equation of (9) for the element of the shrouding. But, as it is very difficult to solve the above equation involving the combined effect of the flexural rigidity and the centrifugal force, I have analysed the problem by separating this in two parts. Thus, to get the resulting frequency equation of the problem containing all effects in the present case, we have to use the equation cited in the introductory part of this paper, namely,

$$p^2 = p'^2 + p''^2. \quad . \quad . \quad . \quad . \quad . \quad (42)$$

The fact that in using the formula not much error is introduced as far as we are concerned with the value of the natural frequency of a rotating elastic body, is evident from the works of many authors.

In order to know the manner of application of the present results, we shall take an example of the following description:

$R = 65{\cdot}3$ cm., $l_1 = 10$ cm., $l_2 = 1$ cm., $a_1 = 0{\cdot}6$ sq. cm., $a_2 = 0{\cdot}3$ sq. cm.;

breadth of blade = effective breadth of shrouding $= 1{\cdot}67$ cm.,

$$\rho_1 = \rho_2 = 6{\cdot}0, \quad E_1 = E_2 = 10^{12}\ \text{C.G.S.}, \quad \omega = 100\,\pi.$$

Using Table I. or fig. 2, we get $\left(\frac{\rho_1 p'^2}{E_1 k_1^2}\right)^{1/4} l_1 = 2{\cdot}235$, from which it follows

$$\frac{p'}{2\pi} = 335 \text{ per sec.}$$

Observing Table II. we obtain $p''^2/\omega^2 = 10$, from which we find

$$\frac{p''}{2\pi} = 158 \text{ per sec.}$$

Hence the actual frequency of vibrations is

$$\frac{p}{2\pi} = \frac{1}{2\pi}\sqrt{p'^2 + p''^2} = 370 \text{ per sec.}$$

When there is no shrouding, we have, from the well-known formula

$$\left(\frac{\rho_1 p'^2}{E_1 k_1^2}\right)^{1/4} l_1 = 1{\cdot}875$$

for a clamped-free bar,

$$\frac{p'}{2\pi} = 237 \text{ per sec.}$$

From fig. 4 we have $p''^2/\omega^2 = 10{\cdot}8$ approximately. Hence

$$\frac{p''}{2\pi} = 165 \text{ per sec.}$$

Therefore the actual frequency of vibrations in this case becomes

$$\frac{p}{2\pi} = \frac{1}{2\pi}\sqrt{p'^2 + p''^2} = 289 \text{ per sec.}$$

XI. *On the Falling of a Strip of Paper.*

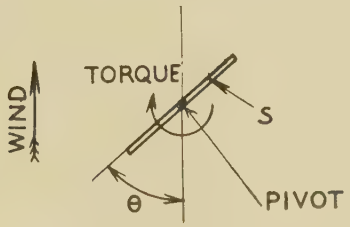
By J. G. Baker, *Pittsburgh, Pa., U.S.A.**

IT is common observation that a strip of paper will spin, when dropped in still air, about an axis parallel to its longest dimension. Furthermore, it will fall in a direction perhaps 40 or 50 degrees from the vertical †.

In order to study this phenomenon, a strip of stiff paper was pivoted on two pins and placed in front of an electric fan. If the strip was given a very small initial rotation it would accelerate and finally rotate steadily at a rather high speed.

When this strip was held at an angle with the air stream, a torque appeared tending to rotate the strip so that its plane would be perpendicular to the direction of the

Fig. 1.



air stream (refer to fig. 1). If the clockwise angle between the wind direction and the stationary strip S is θ, and if $\theta < 90°$, the torque due to the wind on the strip is clockwise.

Four cases can be distinguished, according to the ratio of the velocity of the edge of the strip to the relative wind velocity (no distinction is made between the falling strip and the pivoted one) : (1) in which this ratio is zero, (2) in which it is small, (3) in which the ratio is not negligible with respect to unity, and (4) in which the edge velocity is very high compared with the relative wind.

In (1) it is evident that the torque at an angle θ (fig. 1) is equal and opposite to the torque for angle $-\theta$. The

* Communicated by Prof. R. V. Southwell, M.A., F.R.S.

† This phenomenon was first called to my attention in June 1932, in New Haven, Conn., by Mr. Spencer during the A. S. M. E., Applied Mechanics Division, meeting there.

average torque for a revolution at zero rotational velocities will then be zero. Since the strip accelerates with slow rotation, in case (2) there is an average torque tending to increase the rotation. In case (4) the flow will be about the same as with no relative wind; hence, there will be an average torque in opposition to the rotation. It is apparent that equilibrium is reached in case (3). From here on it will be only attempted to show why there is an average torque in the direction of rotation for case (2).

It is well known that the initial conditions of flow around an object when it is first placed in a wind are widely different from the conditions which obtain when a steady state is reached. Also this difference is greatest in the flow on the down-stream side of the object, due to the accumulation there of turbulent fluid in the steady state *.

On the up-stream side there is practically no change from the first instant until the steady condition is reached. For any change in position of the strip of paper, the up-stream side would not be expected to require appreciable time to reach its new steady state condition. On this account, it may be reasonably assumed that the torque on the up-stream side of the strip of paper (for case 2) depends only on the angle θ,—*i. e.*, is independent of the rotational velocity of the strip. On this assumption, the average torque due to this side is zero.

The flow on the down-stream side of the strip is altered due to the velocity of rotation, because of the time required to accumulate turbulent air back of the strip. It would be expected that this time (for a given angle) would become more important with increase in the difference between :—

A. The change of accumulated turbulent air per unit change in the angle of the strip —considering only the steady state in case (1); and

B. The rate of accumulation of turbulent air if the amount of it is less than that required for the steady state in case (1).

At all events this difference indicates the rapidity with which the steady state is reached. B is larger, the greater the sharpness of turn it would be necessary for the air to

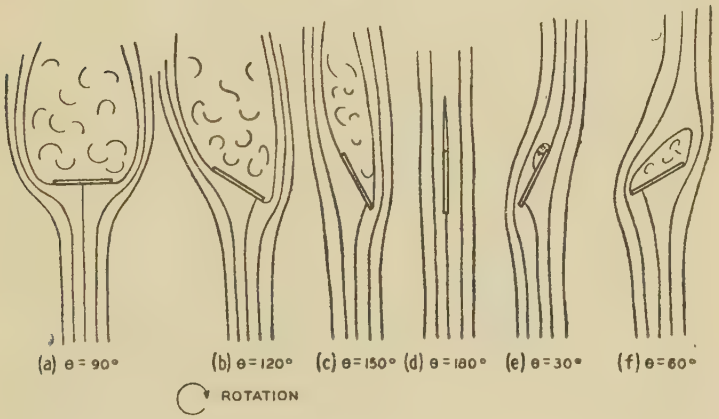
* See 'The Physics of Solids and Fluids,' by Ewald, Pöschl, and Prandtl, pp. 283 and 284, figs. 9 & 10.

make were it to flow along the boundary of the object on the down-stream side.

As an illustration of the effect B, if there is no accumulated turbulent air back of the strip, the time required for the steady state of case (1) to be reached is much longer for $\theta=30°$ than for $\theta=90°$. In fact, for $\theta=90°$ the turn that the air must make to flow along the boundary on back of the strip is so sharp that the necessary turbulent air for the steady state of flow accumulates almost instantaneously.

In fig. 2 is shown the probable general character of the flow around a spinning strip of paper. (The direction of flow is upward. The strip is rotating clockwise.) From the last

Fig. 2.



paragraph we assume that the flow for $\theta=90°$ is the same as the steady state flow with no rotational velocity. For $\theta=120°$ and 150°, etc., the accumulated turbulent air for no rotational velocity is less and less. A surplus of accumulated air is always quickly washed away by direct contact with the main flow. It is to be expected, then, that for the angles in question the flow will be practically the steady flow in spite of the rotation velocity. When $\theta=180°$, the steady state condition is immediately attained, namely, that of no turbulent air. In turning to $\theta=210°$, for the steady state, an accumulation is required ; but the air can easily flow in behind the strip and the accumulation is delayed, and (fig. 2) (*e*) is different from (*c*). Likewise (*f*) is different from (*b*). In (*c*) the

torque due to the down-stream side is nearly zero, because of the pressures being more or less constant in the turbulent-air region. From the experiments with the pivoted strip, it will be remembered that for a stationary strip the torque for $\theta < 90°$ was in a clockwise direction. For these cases, if θ were not too small, the torque must have been entirely due to the front side in view of the above statements.

Now for (*e*) the flow in the rear is not far different from that in front, hence there must be almost twice as much torque as for (*c*) (*c* is the same as the steady state). The net result, if this reasoning is correct, is an average torque in the direction of spin.

The flow in (*e*) and for angles of nearly the same magnitude appears much like that of a good aerofoil, whereas the flow for (*c*) is similar to that round an aerofoil which is inclined beyond the burble point,—*i. e.*, where the lift is small. Due to the spin, therefore, the average resultant force is not in the direction of the wind, but has a component to the left (for clockwise rotation).

For a strip falling at constant velocity, the gravity and the resultant wind force must be equal and opposite, hence the strip in question glides to the left.

This checks experiment—in fact, even though the strip spins so fast that it is case (3), it still descends to the left.

In conclusion, the fact that appreciable time is required for turbulent air to accumulate on the upper side of a falling piece of paper is a likely basis for an explanation of the spinning of such a strip of paper and of its failure to descend vertically in still air.

Research Labs., Westinghouse,
East Pittsburgh, Pa,

XII. *Fermat's Principle.* *By* Dr. Ladislas Natanson, *Professor of Natural Philosophy in the Jagellonian University, Cracow**.

§ 1. FOR the preliminary purpose of introduction, we recapitulate briefly some auxiliary theorems: they will enable us to shorten the calculations. Let x, y, z be

* Communicated by the Author. Papers on Fermat's Principle were read by Professor Lad. Natanson, before the Polish Academy of Sciences and Letters, in 1929, 1930, and 1931, and published in the Academy's *Bulletin International.*

the rectangular coordinates of a point M on a curve L and let l represent the arc measured, along L, from some particular point M_0 of L to the point M contemplated. At the point M we imagine three unit vectors $\mathbf{l}$, $\mathbf{m}$, $\mathbf{n}$ lying in the same relative order as the positive semi-axes of a right-handed system of coordinates x, y, z. We suppose $\mathbf{l}$ localized in the tangent and pointing in the sense in which l increases, $\mathbf{m}$ in the binormal, perpendicular to the osculating plane and $\mathbf{n}$ in the normal, drawn towards the centre of curvature. Then

$$\mathbf{l}={}_{\mathrm{V}}(\mathbf{mn}), \qquad \mathbf{m}={}_{\mathrm{V}}(\mathbf{nl}), \qquad \mathbf{n}={}_{\mathrm{V}}(\mathbf{lm}), \quad . \quad . \quad (1)$$

${}_{\mathrm{V}}(\ )$ being the symbol of the vector product ; and

$$l_x=\frac{dx}{dl}, \qquad l_y=\frac{dy}{dl}, \qquad l_z=\frac{dz}{dl}, \quad . \quad . \quad (2)$$

$$n_x=\mathrm{R}\frac{dl_x}{dl}, \qquad n_y=\mathrm{R}\frac{dl_y}{dl}, \qquad n_z=\mathrm{R}\frac{dl_z}{dl}, \quad . \quad . \quad (3)$$

where R is used to denote the radius of curvature at M. Other expressions of n_x, n_y, n_z may be obtained from the well-known equation

$$\mathbf{n}=-\mathrm{R}_{\mathrm{V}}(\mathbf{l}\,.\,\mathrm{curl}\,\mathbf{l}), \quad . \quad . \quad . \quad . \quad . \quad (4)$$

and for $\mathbf{m}$ we derive

$$\mathbf{m}=\mathrm{R}[{}_{\mathrm{S}}(\mathbf{l}\,.\,\mathrm{curl}\,\mathbf{l})\,.\,\mathbf{l}-\mathrm{curl}\,\mathbf{l}], \quad . \quad . \quad . \quad (5)$$

where the scalar product of two vectors is written ${}_{\mathrm{S}}(\ )$.

In what follows, the operations ${}_{\mathrm{S}}(\mathbf{l}.\nabla)$, ${}_{\mathrm{S}}(\mathbf{m}.\nabla)$ and ${}_{\mathrm{S}}(\mathbf{n}.\nabla)$ will be frequently recurring; we shall represent them respectively by d/dl, d/dm, and d/dn.

§ 2. Let $\mu(x, y, z)$ be used to denote a continuous scalar function of x, y, z, supposed to be given in a domain D in which L is situated. Remembering, then, that $l_x^2+l_y^2+l_z^2=1$, we can easily verify that the following remarkable formula holds :—

$$\nabla\mu-\frac{d}{dl}(\mu\mathbf{l})={}_{\mathrm{V}}(\mathbf{l}\,.\,\mathrm{curl}\,(\mu\mathbf{l})). \quad . \quad . \quad . \quad . \quad (1)$$

In the particular case in which μ is independent of x, y, z, the space variation $\nabla\mu$ vanishes and (1) reduces to

$$-\frac{d\mathbf{l}}{dl}={}_{\mathrm{V}}(\mathbf{l}\,.\,\mathrm{curl}\,\mathbf{l}), \quad . \quad . \quad . \quad . \quad . \quad (2)$$

a simple and easily interpretable equation, frequently employed in Geometrical Optics.

It remains to enunciate our second proposition; for this purpose consider the integral

$$\mathrm{I}=\int_{\mathrm{P}}^{\mathrm{Q}}\mu(x,\,y,\,z)\,dl \quad . \quad . \quad . \quad . \quad . \quad (3)$$

extended, along the curve L, between two given points P and Q. Using this to calculate the variation $\delta\mathrm{I}$ of I we will suppose the initial and the final positions P and Q on L to be invariable; with this understanding, we easily find

$$\delta\mathrm{I}=\int_{\mathrm{P}}^{\mathrm{Q}}\left\{\left(\frac{\partial\mu}{\partial x}-\frac{d\mu l_x}{dl}\right)\delta x+\left(\frac{\partial\mu}{\partial y}-\frac{d\mu l_y}{dl}\right)\delta y\right.$$
$$\left.+\left(\frac{\partial\mu}{\partial z}-\frac{d\mu l_z}{dl}\right)\delta z\right\}dl\,; \quad . \quad (4)$$

hence, employing (1), we obtain

$$\delta\mathrm{I}=\int_{\mathrm{P}}^{\mathrm{Q}}(\mathrm{M}_x\delta x+\mathrm{M}_y\delta y+\mathrm{M}_z\delta z)dl, \quad . \quad . \quad (5)$$

the vector quantity **M** being defined as

$$\mathbf{M}=\frac{1}{\mu}(\mathbf{l}\,.\,\mathrm{curl}\,(\mu\mathbf{l})). \quad . \quad . \quad . \quad . \quad . \quad (6)$$

§ 3. The following is in amplification and development of the results recited above. We consider the line-integral

$$\mathrm{J}=\int_{\mathrm{P}}^{\mathrm{Q}}(\mathrm{A}_x dx+\mathrm{A}_y dy+\mathrm{A}_z dz), \quad . \quad . \quad . \quad . \quad (1)$$

where A_x, A_y, A_z are the components of a vector quantity **A**, and the integration is effected along a curve L extending from a given (invariable) point P to another given (invariable) point Q. From (1), we conclude

$$\delta\mathrm{J}=\int_{\mathrm{P}}^{\mathrm{Q}}\left\{l_x\,.\,\delta\mathrm{A}_x+l_y\,.\,\delta\mathrm{A}_y+l_z\,.\,\delta\mathrm{A}_z\right.$$
$$\left.+\mathrm{A}_x\frac{d\delta x}{dl}+\mathrm{A}_y\frac{d\delta y}{dl}+\mathrm{A}_z\frac{d\delta z}{dl}\right\}dl. \quad . \quad . \quad (2)$$

Integrating by parts, we easily obtain

$$\delta\mathrm{J}=\int_{\mathrm{P}}^{\mathrm{Q}}(\mathrm{N}_x\delta x+\mathrm{N}_y\delta y+\mathrm{N}_z\delta z)dl, \quad . \quad . \quad . \quad (3)$$

where we assume

$$\mathrm{N}_x=l_x\frac{\partial\mathrm{A}_x}{\partial x}+l_y\frac{\partial\mathrm{A}_y}{\partial x}+l_z\frac{\partial\mathrm{A}_z}{\partial x}-\frac{d\mathrm{A}_x}{dl}, \quad . \quad (4a)$$

and two similar definitions for N_y and N_z. Now if we

substitute their full expressions for $d\mathrm{A}_x/dl$, $d\mathrm{A}_y/dl$, and $d\mathrm{A}_z/dl$, equations (4) become

$$\mathrm{N}_x = l_y\left(\frac{\partial \mathrm{A}_y}{\partial x} - \frac{\partial \mathrm{A}_x}{\partial y}\right) - l_z\left(\frac{\partial \mathrm{A}_x}{\partial z} - \frac{\partial \mathrm{A}_z}{\partial x}\right), \quad . \quad (5a)$$

etc., and these declare that N_x, N_y, and N_z are the components of the vector product

$$\mathbf{N} = {}_{\mathrm{V}}(\mathbf{1} \,.\, \mathrm{curl}\, \mathbf{A}). \quad . \quad . \quad . \quad . \quad . \quad . \quad (6)$$

From equation (3) we conclude that $\delta \mathrm{J}$ disappears when $\mathbf{N}$ vanishes, that is, when curl $\mathbf{A}$ either is null or coincides with $\mathbf{1}$ in direction.

§ 4. Our previous results having been brought under a different form of expression, the next enquiry is, how are we to dispose our calculations so as to force them to display their mutual connexion. In the first place we observe that, in the particular case in which

$$\mathbf{A} = \mu(x, y, z)\mathbf{1}, \quad . \quad . \quad . \quad . \quad . \quad . \quad (1)$$

the vector quantity $\mathbf{N}$ defined in (6), § 3, agrees with $\mathbf{M}$ [*cf.* (6), § 2], the integral J degenerates to I [*cf.* (3), § 2], so that equation (3) of § 3 merely reiterates (5), § 2. The theorem established in § 3 does not, however, transcend in generality that contemplated in § 2 ; the integral J can be represented in the form

$$\mathrm{J} = \int_{\mathrm{P}}^{\mathrm{Q}} {}_{\mathrm{S}}(\mathbf{1A}) \,.\, dl, \quad . \quad . \quad . \quad . \quad . \quad . \quad (2)$$

and reduces therefore to the integral I when $\mu(x, y, z)$ of § 2 is restricted to designate the scalar product ${}_{\mathrm{S}}(\mathbf{1A})$.

It will serve to bring out more clearly the nature of the theorems with which we are concerned if we attend to the following process of deduction. We begin by remarking that three equations of the form

$$l_x \delta(dl) = \delta(dx) - \delta l_x \,.\, dl \quad . \quad . \quad . \quad . \quad (3a)$$

evidently are true ; substitute these in the expansion of $\delta(\lambda \,.\, dl)$ where $\lambda = {}_{\mathrm{S}}(\mathbf{1A})$, then we obtain

$$\delta(\lambda \,.\, dl) = \delta\lambda \,.\, dl + \mathrm{A}_x\delta(dx) + \mathrm{A}_y\delta(dy) + \mathrm{A}_z\delta(dz) - (\mathrm{A}_x \,.\, \delta l_x + \mathrm{A}_y \,.\, \delta l_y + \mathrm{A}_z \,.\, \delta l_z)dl. \quad . \quad (4)$$

Proceeding as before, we get the result

$$\int_P^Q \left\{\left(\frac{\partial\lambda}{\partial x}-\frac{d.\lambda l_x}{dl}\right)\delta x+\left(\frac{\partial\lambda}{\partial y}-\frac{d.\lambda l_y}{dl}\right)\delta y+\left(\frac{\partial\lambda}{\partial z}-\frac{d.\lambda l_z}{dl}\right)\delta z\right\}dl$$
$$=\int_P^Q [\delta\lambda-(A_x\delta l_x+A_y\delta l_y+A_z\delta l_z)]dl$$
$$-\int_P^Q (dA_x.\delta x+dA_y.\delta y+dA_z.\delta z). \quad (5)$$

Consequently, if we put

$$\mathbf{M}={}_V(1.\operatorname{curl}(\lambda 1)), \text{ where } \lambda={}_S(1A)\ ;\ \text{and } \mathbf{N}={}_V(1.\operatorname{curl}\mathbf{A}), \quad (6)$$

we ultimately have the equality

$$\int_P^Q (M_x\delta x+M_y\delta y+M_z\delta z)dl=\int_P^Q (N_x\delta x+N_y\delta y+N_z\delta z)dl \quad (7)$$

which answers our enquiry; the desired connexion is thus established. To elucidate the nature of this equality, let us consider the vector difference

$$\mathbf{D}=\lambda 1-\mathbf{A}. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (8)$$

If we write for the moment

$$\Pi={}_V(1.\operatorname{curl}\mathbf{D}), \quad . \quad . \quad . \quad . \quad . \quad . \quad (9)$$

we have, by what has gone before,

$$\int_P^Q (\Pi_x\delta x+\Pi_y\delta y+\Pi_z\delta z)dl=0. \quad . \quad . \quad . \quad (10)$$

By application of the theorem of § 3, equation (10) reduces itself to

$$\delta\int_P^Q {}_S(1\mathbf{D}).dl=0. \quad . \quad . \quad . \quad . \quad . \quad (11)$$

Since however, we evidently have, by (8), ${}_S(1\mathbf{D})=0$, it follows that (11), and consequently (10) and (7), are identical propositions.

§ 5. Considering again, as in § 3, the line integral J, we proceed to calculate its variation δJ and easily find

$$\delta J=\int_P^Q [\delta A_x.dx+\delta A_y.dy+\delta A_z.dz$$
$$-(dA_x.\delta x+dA_y.\delta y+dA_z.\delta z)]. \quad . \quad (1)$$

For δA_x, δA_y, δA_z, dA_x, dA_y, dA_z we substitute their full expressions; equation (1) becomes

$$\delta J=\int_P^Q [(l_y\operatorname{curl}_z\mathbf{A}-l_z\operatorname{curl}_y\mathbf{A})\delta x$$
$$+(l_z\operatorname{curl}_x\mathbf{A}-l_x\operatorname{curl}_z\mathbf{A})\delta y$$
$$+(l_x\operatorname{curl}_y\mathbf{A}-l_y\operatorname{curl}_x\mathbf{A})\delta z]dl, \quad . \quad . \quad (2)$$

and this reiterates the results given above and expressed by (3), (4), (5), and (6), § 3.

The integration in J is effected along a curve L extending from a given point P to another given point Q. Now let another curve Λ intersect L in (x, y, z). Conceive a portion of surface S and let its boundary be supposed to consist of a single closed curve O; S is so described as to enclose two sets of lines, one set of L-lines and another of Λ-lines; the points P and Q (between which the L-lines are laid) are situated on the boundary O of S. We use dx, dy, dz to denote the components of an element $\mathbf{dl}$ of L and likewise δx, δy, δz to denote the components of an element $\delta\mathbf{l}$ of Λ. Writing

$$\mathbf{N}.d\mathrm{S} = {}_{\mathrm{V}}(\mathbf{dl}.\delta\mathbf{l}), \quad . \quad . \quad . \quad . \quad . \quad (3)$$

we know that $\mathbf{N}$, a unit vector normal to dS, is always drawn on one side of S, which we may term the positive side.

Starting from these assumptions, we consider, as above, the line integral J, taken along a L-line; the variation δJ which arises as we proceed from a given L-line to a neighbouring L-line, adjacent on the surface, is, by (2),

$$\begin{aligned}\delta\mathrm{J} = -\int_{\mathrm{P}}^{\mathrm{Q}} \{&(dy.\delta z - dz.\delta y)\,\mathrm{curl}_x\,\mathbf{A} \\ &+ (dz.\delta x - dx.\delta z)\,\mathrm{curl}_y\,\mathbf{A} \\ &+ (dx.\delta y - dy.\delta x)\,\mathrm{curl}_z\,\mathbf{A}\}, \quad . \quad . \quad (4)\end{aligned}$$

or, which is the same thing,

$$\delta\mathrm{J} = -\int_{\mathrm{P}}^{\mathrm{Q}} d\mathrm{S}.{}_{\mathrm{S}}(\mathbf{N}.\mathrm{curl}\,\mathbf{A}). \quad . \quad . \quad . \quad . \quad . \quad (5)$$

Now let us select for examination an element PQ of S enclosed between contiguous L-lines; clearly, $-\delta$J is the circulation (say J_{PQ}) of $\mathbf{A}$ round the boundary of PQ, provided that circulation be taken in the sense positively related to the direction of $\mathbf{N}$. By a well-known proposition, the sum, extended to the surface S, of circulations J_{PQ} following the course of the boundaries of all strips PQ, is equal in amount to J_0, the integration in J_0 being performed round the circuit O; thus

$$\mathrm{J}_0 = \sum_{\mathrm{S}} \mathrm{J}_{\mathrm{PQ}} = -\sum_{\mathrm{S}} \delta\mathrm{J} = \iint_{\mathrm{S}} d\mathrm{S}.{}_{\mathrm{S}}(\mathbf{N}.\mathrm{curl}\,\mathbf{A}). \quad . \quad (6)$$

Although first explicitly enunciated by Lord Kelvin (*cf.* 'Mathematical and Physical Papers' by the late Sir G. G. Stokes, Bart., vol. v. Cambridge, 1905, pp. 320–321. Note by Sir Joseph Larmor), this result has been known ever since as *Stokes' Theorem*.

§ 6. Let us suppose that **A** coincides with **l** in direction so that the equations of L are

$$dx : dy : dz = l_x : l_y : l_z = A_x : A_y : A_z. \quad . \quad . \quad (1)$$

On this understanding we proceed : in attempting to apply our scheme of equations to classes of widely different cases we shall find the present simplifying assumption appropriate to our purpose.

For convenience, let us employ the symbol $|\,A\,|$ to denote the absolute value of **A** :

$$|\,A\,| = \text{abs. val. of } \sqrt{(A_x^2 + A_y^2 + A_z^2)}\ ; \quad . \quad . \quad (2)$$

then the following admits of easy verification :

$$\frac{\partial}{\partial x}\,|\,A\,| = l_x \frac{\partial A_x}{\partial x} + l_y \frac{\partial A_y}{\partial x} + l_z \frac{\partial A_z}{\partial x}, \text{ etc.} \quad . \quad (3a)$$

Equations (4), § 3, reduce accordingly to

$$N_x = \frac{\partial}{\partial x}\,|\,A\,| - \frac{dA_x}{dl}, \text{ etc.}\ ; \quad . \quad . \quad . \quad (4a)$$

the line integral J defined in (1), § 3, can now be written in the form

$$J = \int_P^Q |\,A\,| \,.\, dl. \quad . \quad . \quad . \quad . \quad . \quad (5)$$

So much being granted, we infer from (6), § 3, and (3), § 1, that the relation

$$\nabla\,|\,A\,| = \mathbf{l}\frac{d\,|\,A\,|}{dl} + \mathbf{n}\frac{|\,A\,|}{R} + {}_v(\mathbf{l}\,.\,\mathrm{curl}\,\mathbf{A}) \quad . \quad (6)$$

holds good. This equation, to which we shall have frequent recourse, may be established by a different process. We begin by observing that, for any continuous scalar function $|\,A\,|$ of the coordinates, the identity is true :

$${}_v(\mathbf{l}\,.\,{}_v(\mathbf{l}\,.\,\nabla\,|\,A\,|)) = -\nabla\,|\,A\,| + \mathbf{l}\frac{d\,|\,A\,|}{dl}. \quad . \quad (7)$$

And again curl **A** or curl $|\,A\,|\,\mathbf{l}$ is expressible in the form

$$\mathrm{curl}\,|\,A\,|\,\mathbf{l} = |\,A\,|\,\mathrm{curl}\,\mathbf{l} - {}_v(\mathbf{l}\,.\,\nabla\,|\,A\,|), \quad . \quad (8)$$

so that, by application of (4), § 1, we are led to the relation

$${}_v(\mathbf{l}\,.\,\mathrm{curl}\,\mathbf{A}) = -\mathbf{n}\,\frac{|\,A\,|}{R} - {}_v(\mathbf{l}\,.\,{}_v(\mathbf{l}\,.\,\nabla\,|\,A\,|))\,; \quad . \quad (9)$$

combining this with (7) we are brought back to the foregoing equation (6).

A number of interesting results are connected with equation (6). By applying to both sides the operations $_s(\mathbf{m}.)$ and $_s(\mathbf{n}.)$, it appears that we have

$$\frac{d\,|\,\mathrm{A}\,|}{dm} = -{}_s(\mathbf{n}\,.\,\mathrm{curl}\,\mathbf{A}), \quad . \quad . \quad . \quad . \quad (10)$$

$$\frac{d\,|\,\mathrm{A}\,|}{dn} = \frac{|\,\mathrm{A}\,|}{\mathrm{R}} + {}_s(\mathbf{m}\,.\,\mathrm{curl}\,\mathbf{A}). \quad . \quad (11)$$

To arrive at (10) by another mode of procedure, it will be convenient to make use anew of (4), § 1, from which we conclude that $_s(\mathbf{n}\,.\,\mathrm{curl}\,\mathbf{l})=0$. Taken conjointly with this, equation (8) above gives

$${}_s(\mathbf{n}\,.\,\mathrm{curl}\,|\,\mathrm{A}\,|\,\mathbf{l}) = -{}_s(\mathbf{n}\,.{}_v(\mathbf{l}\,.\,\nabla\,|\,\mathrm{A}\,|)) = -{}_s(\mathbf{m}\,.\,\nabla\,|\,\mathrm{A}\,|), \quad . \quad . \quad . \quad (12)$$

which is in conformity with (10). The analogous formula (11) is obtained in a similar way. Returning to (4) and (5) of § 1 it is easily found that $_s(\mathbf{m}\,.\,\mathrm{curl}\,\mathbf{l}) = -1/\mathrm{R}$; now let us refer again to (8) and we shall readily see that

$$\begin{aligned} {}_s(\mathbf{m}\,.\,\mathrm{curl}\,|\,\mathrm{A}\,|\,\mathbf{l}) &= -\frac{|\,\mathrm{A}\,|}{\mathrm{R}} - {}_s(\mathbf{m}\,.{}_v(\mathbf{l}\,.\,\nabla\,|\,\mathrm{A}\,|)) \\ &= -\frac{|\,\mathrm{A}\,|}{\mathrm{R}} + {}_s(\mathbf{n}\,.\,\nabla\,|\,\mathrm{A}\,|); \quad . \quad . \quad (13) \end{aligned}$$

this is essentially equation (11) above.

§ 7. We shall not stay to discuss any of the special results obtained but proceed to consider examples of their application. If we suppose that the vector **A** has no curl, equation (6), § 6, is simplified, and becomes

$$\nabla\,|\,\mathrm{A}\,| = \mathbf{l}\frac{d\,|\,\mathrm{A}\,|}{dl} + \mathbf{n}\,\frac{|\,\mathrm{A}\,|}{\mathrm{R}}. \quad . \quad . \quad . \quad (1)$$

This may be otherwise expressed by saying that

$${}_s(\mathbf{m}\,.\,\nabla\,|\,\mathrm{A}\,|) = \frac{d\,|\,\mathrm{A}\,|}{dm} = 0, \quad . \quad . \quad . \quad . \quad (2)$$

$${}_s(\mathbf{n}\,.\,\nabla\,|\,\mathrm{A}\,|) = \frac{d\,|\,\mathrm{A}\,|}{dn} = \frac{|\,\mathrm{A}\,|}{\mathrm{R}}. \quad . \quad (3)$$

Whenever our present assumption is legitimate and (1)

accordingly may be accepted, other useful formulæ can be obtained ; for instance,

$$_{\mathrm{V}}(\mathbf{l}\,.\,\nabla\,|\,\mathrm{A}\,|\,)=-\mathrm{m}\frac{|\,\mathrm{A}\,|}{\mathrm{R}},\quad .\quad .\quad .\quad .\quad (4)$$

$$_{\mathrm{V}}(\mathbf{n}\,.\nabla\,|\,\mathrm{A}\,|\,)=+\mathrm{m}\frac{d\,|\,\mathrm{A}\,|}{dl}.\quad .\quad .\quad .\quad (5)$$

If we look at equations (3) and (6) of § 3, we may see our way to an important conclusion. So long as curl **A** vanishes we know that $\mathbf{N}=0$, whence it follows that, under such restriction as this, $\delta\mathrm{J}=0$. By $\mathbf{N}=0$ used in equations (4), § 6, we find under these circumstances

$$\frac{\partial}{\partial x}\,|\,\mathrm{A}\,|=\frac{d\mathrm{A}_x}{dl},\quad \frac{\partial}{\partial y}\,|\,\mathrm{A}\,|=\frac{d\mathrm{A}_y}{dl},\quad \frac{\partial}{\partial z}\,|\,\mathrm{A}\,|=\frac{d\mathrm{A}_z}{dl}.\ (6)$$

§ 8. Up to this point it will be observed that we have abstained from attributing any particular physical meaning to the quantities involved in our scheme of interlaced connexions. The interest of that abstract apparatus follows, however, from its application to natural phenomena ; and the point to which attention is here directed is that its field can be found amongst various provinces of Physical Science.

Turning first to Geometrical Optics, let us consider a non-homogeneous medium transmitting luminous disturbances along curvilinear paths L of propagation, the unit vector l pointing in the direction of propagation. We assume, as usual, that every ray L is orthogonal everywhere to surfaces of constant phase

$$\mathrm{Q}(x,\,y,\,z,\,t)=\text{const.}\quad .\quad .\quad .\quad .\quad .\quad (1)$$

In order to investigate the rate of motion of one of these surfaces, normal to itself, let us imagine an observer running his eye along a ray L with velocity **V** ; in time dt the observer perceives the following change in the value of Q :

$$dQ=dt\left\{{}_{\mathrm{s}}(\mathbf{V}\,.\,\nabla\mathrm{Q})+\frac{\partial\mathrm{Q}}{\partial t}\right\}.\quad .\quad .\quad .\quad .\quad (2)$$

A vector **V** localized in the tangent to the ray is said to represent the velocity of phase propagation if it satisfies the condition

$$d\mathrm{Q}=0\quad \text{or}\quad {}_{\mathrm{s}}(\mathbf{V}\,.\,\nabla\mathrm{Q})=-\frac{\partial\mathrm{Q}}{\partial t}.\quad .\quad .\quad .\quad (3)$$

To show the application of (3), we introduce a unit vector **q** localized in the tangent to L and pointing in the direction

of ∇Q. By this definition ${}_s(\mathrm{lq}) = \pm 1$, and from (3) we obtain

$$| \mathrm{V} | = - \frac{1}{{}_s(\mathrm{lq}) \cdot | \nabla Q |} \cdot \frac{\partial Q}{\partial t} \quad . \quad . \quad . \quad . \quad (4)$$

Looking to (4), we see that two cases can occur:

$$\frac{\partial Q}{\partial t} < 0 : {}_s(\mathrm{lq}) = +1 \quad \text{and} \quad \mathrm{l} = +\mathrm{q}, \quad . \quad . \quad (5a)$$

$$\frac{\partial Q}{\partial t} > 0 : {}_s(\mathrm{lq}) = -1 \quad \text{and} \quad \mathrm{l} = -\mathrm{q}. \quad . \quad . \quad (5b)$$

In both cases equation (4) may be written in the form of Sir W. R. Hamilton's fundamental relation

$$\frac{1}{\mathrm{V}^2} \left(\frac{\partial Q}{\partial t} \right)^2 = \left(\frac{\partial Q}{\partial x} \right)^2 + \left(\frac{\partial Q}{\partial y} \right)^2 + \left(\frac{\partial Q}{\partial z} \right)^2 . \quad . \quad (6)$$

As may be inferred from equation (4), the velocity $| \mathrm{V} |$ of phase propagation depends in general upon x, y, z, and t. Assuming that this is so, we can define the refractive index, at a point (x, y, z) and time t, to be

$$\mu = \frac{c}{| \mathrm{V} |} = \mu(x, y, z, t), \quad . \quad . \quad . \quad . \quad (7)$$

c representing, as usual, the wave velocity in free empty space; and we introduce what may perhaps be named *Fermat's Vector*:

$$\mu(x, y, z, t) \cdot \mathrm{l} = -c \frac{\nabla Q}{\frac{\partial Q}{\partial t}} . \quad . \quad . \quad . \quad . \quad (8)$$

If we write

$$\mu \mathrm{l} = \theta(x, y, z, t) \cdot \nabla Q, \quad . \quad . \quad . \quad . \quad (9)$$

on the understanding that

$$\theta(x, y, z, t) \frac{\partial Q}{\partial t} = -c, \quad . \quad . \quad . \quad . \quad (10)$$

it easily appears that

$$\mathrm{curl}\,(\mu \mathrm{l}) = {}_v(\nabla \theta \cdot \nabla Q) = \frac{c}{\left(\frac{\partial Q}{\partial t} \right)^2} {}_v \left(\nabla \left(\frac{\partial Q}{\partial t} \right) \cdot \nabla Q \right). \quad (11)$$

We now require the expression of the vector product

$$\mathbf{M} = {}_v(\mathrm{l} \,.\, \mathrm{curl}\,(\mu \mathrm{l})). \quad . \quad . \quad . \quad . \quad (12)$$

Its x-component is

$$M_x = \frac{\pm c}{\left(\frac{\partial Q}{\partial t}\right)^2}\left\{ \mid \nabla Q \mid \frac{\partial}{\partial t}\left(\frac{\partial Q}{\partial x}\right) - \frac{\partial Q}{\partial x}\frac{\partial}{\partial t} \mid \nabla Q \mid \right\}$$

$$= \frac{c}{\left(\frac{\partial Q}{\partial t}\right)^2} \mid \nabla Q \mid^2 \frac{\partial l_x}{\partial t}$$

$$= \frac{\mu^2}{c} \cdot \frac{\partial l_x}{\partial t}. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (13\,a)$$

From (13 a) and two similar equations we learn that

$$\mathbf{M} = \frac{\mu^2}{c} \cdot \frac{\partial \mathbf{l}}{\partial t}, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (14)$$

and have thus verified the relation [*cf.* (1), § 2 above]:

$$\nabla\mu - \frac{d}{dl}(\mu\mathbf{l}) = \frac{\mu^2}{c} \cdot \frac{\partial \mathbf{l}}{\partial t}. \quad . \quad . \quad . \quad . \quad . \quad (15)$$

§ 9. From equation (11) in the foregoing § 8 we infer that curl $(\mu\mathbf{l})$ is at right angles to the direction of $\mathbf{l}$. The same conclusion follows by observing that

$$\mathrm{curl}_x(\mu\mathbf{l}) = -\frac{c\,\frac{\partial Q}{\partial y} \cdot \frac{\partial Q}{\partial z}}{\left(\frac{\partial Q}{\partial t}\right)^2} \cdot \frac{\partial}{\partial t} \log \mathrm{nat} \frac{\frac{\partial Q}{\partial y}}{\frac{\partial Q}{\partial z}}, \text{ etc.,} \quad . \quad (1\,a)$$

whence

$$_s(\mathbf{l} \,.\, \mathrm{curl}\,(\mu\mathbf{l})) = 0. \quad . \quad . \quad . \quad . \quad . \quad (2)$$

In Geometrical Optics it is usual to suppose that everything in the medium has settled into a steady state; when such a permanent collocation is attained, the wave-disturbances are capable of maintaining themselves indefinitely without readjustment; in a stationary state, therefore, the differential coefficients $\partial l_x/\partial t$, $\partial l_y/\partial t$, $\partial l_z/\partial t$ must vanish, so that M_x, M_y, M_z, and consequently $\mathbf{M}$, cannot differ from zero. According to our present result (2), however, curl $(\mu\mathbf{l})$ cannot coincide with $\mathbf{l}$ in direction; thus, finally, for a steady field of radiation we must have

$$\mathrm{curl}\,(\mu\mathbf{l}) = 0, \quad . \quad . \quad . \quad . \quad . \quad . \quad (3)$$

and this is another way of telling us that

$$\mu\mathbf{l} = \nabla \mathrm{E}(x, y, z, t), \quad . \quad . \quad . \quad . \quad . \quad (4)$$

the symbol $\mathrm{E}(x, y, z, t)$ being used to denote a continuous scalar function of the coordinates and the time, known as the *Eiconal Function.* Whenever, therefore, an $\mathrm{E}(x, y, z, t)$ exists, the following relations hold good :

$$\mu^2 = |\nabla \mathrm{E}|^2 \ldots ; \quad . \quad . \quad . \quad . \quad . \quad . \quad (5)$$

$$l_x = \frac{\pm \dfrac{\partial \mathrm{E}}{\partial x}}{|\nabla \mathrm{E}|}, \text{ etc.} \ldots . \quad . \quad . \quad . \quad . \quad . \quad (6\,a)$$

showing that the Eiconal Function is suitable for exploring permanent fields of radiation.

The object of the above reasoning was to show that the vanishing of curl $(\mu \mathbf{l})$, and the existence of the Eiconal Function, requires $\partial \mathbf{l}/\partial t$ to be null ; the reciprocal statement, of course, is also true. These conclusions may be tested by observing that

$$(\text{curl } (\mu \mathbf{l}))^2 = \frac{\mu^4}{c^2} \left\{ \left(\frac{\partial l_x}{\partial t}\right)^2 + \left(\frac{\partial l_y}{\partial t}\right)^2 + \left(\frac{\partial l_z}{\partial t}\right)^2 \right\} \quad . \quad (7)$$

which, by virtue of (2), is compatible with (14), § 8, and proves the theorem.

The remarkable effect of the Eiconal Vector Theorem (4) is to show with full expressiveness how, from the scheme of abstract relations discussed in §§ 2–7, the fundamental statement emerges which is known as Fermat's Principle in Geometrical Optics. It follows indeed from (5), § 2, that when we restrict the calculation to the limiting case of a steady field of radiation (so that $\mathbf{M} = 0$) we immediately find $\delta \mathrm{I} = 0$; and again from (1), § 2, we derive, on the same understanding,

$$\nabla \mu - \frac{d}{dl}(\mu \mathbf{l}) = 0 \quad . \quad . \quad . \quad . \quad . \quad . \quad (8)$$

This differential vector equation embodies Fermat's Principle embracing, in a condensed kinematical formulation, all essential features of experience in the domain of Geometrical Optics.

§ 10. The best example that we can give in illustration of the preceding discussion is afforded by assuming

$$\mathrm{Q}(x, y, z, t) = n\{t - \mathbf{R}(x, y, z)\}, \quad . \quad . \quad . \quad (1)$$

where n (or $2\pi \times$ frequency) is a positive constant, $\mathbf{R}(x, y, z)$

a scalar function of x, y, z which does not contain t explicitly. Whenever the assumption (1) is appropriate, we have

$$\frac{\partial Q}{\partial t} = n > 0\ ;\ \nabla Q = -n\,.\,\nabla \mathbf{R}\ ;\ n\theta = -c, \quad . \quad . \quad (2)$$

and from (9), § 8, the conclusion follows :

$$\mu \mathbf{l} = c\,.\,\nabla \mathbf{R}. \quad . \quad . \quad . \quad . \quad . \quad . \quad (3)$$

By application of (5 *b*), § 8, $\mathbf{l} = -\mathbf{q}$, so that $\mathbf{l}$ is a unit vector complying with $\nabla \mathbf{R}$ in direction. The vector equation (3) gives the scalar result

$$\mu = c \mid \nabla \mathbf{R} \mid, \quad . \quad . \quad . \quad . \quad . \quad . \quad (4)$$

reverting thus, with $\mathrm{E} = c\mathbf{R}$, to the scalar aspect of the Eiconal Theorem (§ 9) ; while

$$\mu \mathbf{l} = \nabla \mathrm{E} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (5)$$

asserts, furthermore, that the direction of $\nabla \mathrm{E}$ coincides with the direction of $\mathbf{l}$. We learn, in addition, that, in the particular case we are contemplating, μ and V are independent of t.

Considering x, y, z as functions of l, we easily find

$$\mathbf{R}(l) = \int \frac{dl}{\mathrm{V}(l)} \quad \text{and} \quad \mathrm{E}(l) = \int \mu(l)\, dl \quad . \quad (6)$$

the Eiconal agrees thus exactly with Fermat's integral.

Bearing in mind that we are now concerned with the particular case in which the paths of propagation of light are invariable and the refractive index of the medium is independent of t, we define as follows the vector $\mathbf{A}$ contemplated in § 6 :

$$\mathbf{A} = \mu(x, y, z)\,.\,\mathbf{l}, \quad . \quad . \quad . \quad . \quad . \quad . \quad (7)$$

and from (5) we conclude : $\mathrm{curl}\,\mathbf{A} = 0$. We are thus justified in asserting that

$$\nabla \mu = \mathbf{l} \frac{d\mu}{dl} + \mathbf{n} \frac{\mu}{\mathrm{R}}, \quad . \quad . \quad . \quad . \quad . \quad . \quad (8)$$

$$\frac{d\mu}{dm} = 0, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (9)$$

and

$$\frac{d\mu}{dn} = \frac{\mu}{\mathrm{R}}\ ; \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (10)$$

cf. (6), (10), and (11), § 6 ; also (1), (2), (3), § 7. The very simple and elegant theorem embodied in equation (10) is due to James Thomson * ; according to P. G. Tait, it was already involved in one of the propositions of W. H. Wollaston's memoir published in 1800. In a paper "On Mirage" †, and again in the delightful elementary volume on 'Light' ‡, P. G. Tait shows that James Thomson's equation immediately follows from the principles of the Corpuscular Theory of Light and adduces examples of problems where it is usefully applicable.

§ 11. We can quite as readily apply the same general formulæ to other appropriate cases. If E_x, E_y, E_z be the components of the intensity, at a point (x, y, z), in an electrostatic field of force, the lines of force defined to satisfy

$$dx : dy : dz = l_x : l_y : l_z = E_x : E_y : E_z, \quad . \quad . \quad . \quad (1)$$

may be identified with the curves L considered in our introductory sections. And since, in an electrostatic field of force, curl $\mathbf{E}=0$, the results of § 7 may be employed. We have at once

$$\frac{\partial}{\partial x} \mid E \mid = \frac{dE_x}{dl}, \quad \frac{\partial}{\partial y} \mid E \mid = \frac{dE_y}{dl}, \quad \frac{\partial}{\partial z} \mid E \mid = \frac{dE_z}{dl}, \quad (2)$$

where $\mid E \mid$ represents the absolute value of $\mathbf{E}$. These are equivalent to $\mathbf{N}=0$, and

$$\delta J = 0, \quad \text{where} \quad J = \int_P^Q \mid E \mid dl. \quad . \quad . \quad . \quad (3)$$

It is also true that

$$\nabla \mid E \mid = \mathbf{l} \frac{d \mid E \mid}{dl} + \mathbf{n} \frac{\mid E \mid}{R}, \quad . \quad . \quad . \quad (4)$$

from which, as before, we deduce

$$\frac{d \mid E \mid}{dm} = 0 ; \quad \frac{d \mid E \mid}{dn} = \frac{\mid E \mid}{R}. \quad . \quad . \quad . \quad (5)$$

This is nothing else than the translation, into the language of Electrostatics, of the principle underlying James

* 'Report of the British Association' (A), Brighton, 1872, p. 41 ; 'Collected Papers in Physics and Engineering,' by James Thomson, selected and arranged by Sir J. Larmor and J. Thomson, Cambridge University Press, p. 441 (1912); *cf.* J. D. Everett's paper "On the Optics of Mirage," Phil. Mag. (4) xlv. pp. 161 and 248 (1873).

† Trans. Roy. Soc. of Ed., xxx., read December 5, 1881 ; 'Scientific Papers,' by P. G. Tait, i., Cambridge University Press, 1898, p. 446, footnote.

‡ 'Light,' by P. G. Tait, 2nd ed., Edinburgh, 1889, pp. 150–151, § 183.

Thomson's theorem. By properly choosing the respective vectors, we are enabled to make lines of force, in an electrostatic field, analogous to rays transmitted, in a stationary optical field, by a non-homogeneous medium ; and we find reproduced in them the same geometrical characteristics.

§ 12. An additional example may be offered for illustration. Consider a material point, of mass **m**, describing a curve L in a conservative field of force. The velocity of motion, at a given point (x, y, z), being $v(x, y, z)$, we adopt $\mathbf{m}v\mathrm{l}$ for our foregoing **A**. Recur to the considerations of § 7 above. If we assume that curl $(\mathbf{m}v\mathrm{l})$ vanishes, an immediate consequence are the equations

$$\frac{\partial \mathbf{m}v}{\partial x} = \frac{d\mathbf{m}v_x}{dl}, \quad \frac{\partial \mathbf{m}v}{\partial y} = \frac{d\mathbf{m}v_y}{dl}, \quad \frac{\partial \mathbf{m}v}{\partial z} = \frac{d\mathbf{m}v_z}{dl} \qquad (1)$$

which, taken conjointly with the assumed principle of conservation of energy, are the equations of motion of the particle **m**. We have accordingly

$$\frac{d\mathbf{m}v}{dm} = 0 \quad \text{and} \quad \frac{d\mathbf{m}v}{dn} = \frac{\mathbf{m}v}{\mathrm{R}}, \quad . \quad . \quad . \quad (2)$$

and these formulæ bring into prominence another aspect of the motion. To appreciate them, we need only introduce the force **F** sustained by the particle. Then

$$\mathrm{F}_x = v\frac{\partial \mathbf{m}v}{\partial x}, \quad \mathrm{F}_y = v\frac{\partial \mathbf{m}v}{\partial y}, \quad \mathrm{F}_z = v\frac{\partial \mathbf{m}v}{\partial z}, \quad . \quad . \quad (3)$$

and we obtain the familiar results

$$\mathrm{F} = v\frac{d\mathbf{m}v}{dl}, \qquad \mathrm{F}_m = 0, \qquad \mathrm{F}_n = v\frac{\mathbf{m}v}{\mathrm{R}}, \quad . \quad . \quad (4)$$

the last of which is the counterpart of James Thomson's optical equation. We can thus explain why P. G. Tait found it easy to derive that equation from the assumptions of the Corpuscular Theory.

An effective way of tracing principles of permanent order in the course of natural phenomena lies in the development of the consequences of that similarity of quantitative relations which is the essence of exact analogy. Specific constructive hypotheses may be successful as instruments for progress ; they may be resorted to for vividness of picture and clearness of description ; still we cannot regard them as being anything else than examples of artificial imagery, illustrating the tendency of the mind to objectify abstract conceptions.

XIII. *High Frequency Discharges in Argon.*—Part II. *Spectral Properties.* *By* S. P. McCallum, *M.A., D.Phil., Fellow of New College, Oxford,* L. Klatzow, *B.Sc., Lincoln College, Oxford, and* J. E. Keyston, *Magdalen College, Oxford.**

[Plate V.]

IN the course of the investigations of the electrical properties of argon described in part I. of this paper † it was found necessary to make a careful spectroscopic test of the purity of each sample of the gas. This investigation revealed many interesting features of the argon spectrum, the most important of which have been mentioned in a preliminary note ‡. In this paper a more complete description of the investigation is given.

One of the most interesting features of the radiation from discharges in pure argon is the appearance of a continuous spectrum stretching from the wave-length 6100 Å. to 2200 Å., which is the shortest wave-length attainable with the Bellingham and Stanley No. 2 spectrograph used in this investigation.

The continuous spectrum is observed in both the striated and unstriated forms of the high-frequency discharge, and in both forms its intensity, as compared with that of the line spectrum, increases when the pressure of the gas is increased. There are two forms of line spectrum obtainable in an argon discharge, known as the red spectrum and the blue spectrum, and in the striated form of the discharge the lines belonging to the red spectrum are clearly seen superimposed upon the continuous background; but in the uniform unstriated discharge only a few faint lines, which also belong to the red spectrum, stand out above the continuous spectrum. At the highest pressure attained, 150 mm., only the one line of wave-length 6965 Å. could be observed. The blue spectrum is visible when the pressure of the gas is very low, less than one millimetre of mercury, in which circumstances the continuous spectrum is absent. Photographs of the spectra for different pressures, from 150 mm. to 0·5 mm. pressure, are given in Pl. V.

* Communicated by Prof. J. S. Townsend, F.R.S.

† McCallum and Klatzow, Phil. Mag. xv. p. 829 (1933).

‡ McCallum, Klatzow, and Keyston, 'Nature,' cxxx. p. 110 (1932).

It is impossible to explain the continuous spectrum on the hypothesis that it is due to recombination of positive ions with electrons, and it is difficult to maintain that it can be due to molecules of the gas. If it were caused by the recombination of electrons with positive ions of the gas it should appear only on the short-wave side of the head of the line series limits. That is, its long-wave limit should be at about 3500 Å. Furthermore, a recombination spectrum should be accompanied by the appearance of very high members of the line series with considerable intensity. Actually the reverse of this is the case, for at the highest pressures even the lines $1s_n$ ($n=2$, 3, 4, 5)$-3p_n$ ($n=1$ to 10) (Megger's notation), which are the second members of the line series, are missing.

In the case of argon it is difficult to attribute the continuous spectrum to molecules of the gas, as is usual in the case of hydrogen, for argon has no known band spectrum, and presumably no argon molecule exists.

Some conclusions concerning the excitation potential of the continuous spectrum can, however, be drawn from a consideration of the electric forces in the discharge, which are given in Part 1 of this paper. In investigations of the motion of electrons in gases* it was found that the mean energy of agitation, $\bar{E}$, of the electrons depends upon the ratio of the electric force, Z, to the pressure, p, of the gas. In argon at 40 mm. pressure the ratio Z/p was found to be 0·70, and the corresponding value of $\bar{E}$ is 9 volts. When the pressure is 150 mm. the value of Z/p is 0·37 and the value of $\bar{E}$ is 6 volts. Thus as $\bar{E}$ diminishes, the intensity of the lines diminishes in comparison with that of the continuous spectrum, and from general considerations of the effects depending upon the distribution of the energies of the electrons it follows that the continuous spectrum is excited when the energy of the electrons is less than that required to excite the line spectrum. The decrease in the mean energy of agitation of the electrons from 9 volts to 6 volts as the pressure is increased from 40 mm. to 150 mm., accompanied by an increase in the intensity of the continuous spectrum, seems to indicate that the excitation potential of this spectrum is not higher than 7 or 8 volts, and may even be considerably less.

* 'The Motion of Electrons in Gases' (Clarendon Press, Oxford, 1925).

The changes in the relative intensities of the continuous and line spectra with the pressure have also been observed in helium* and neon†. The continuous spectrum in argon has many properties similar to those observed in helium and neon. In both these gases the intensity of the continuous spectrum increases in comparison with that of the line spectrum as the mean energy of agitation of the electrons diminishes, so that for both these gases it seems necessary to conclude that the excitation potentials of the continuous spectra are lower than the excitation potentials of the line spectra. A continuous spectrum has also been observed by Meggers in krypton ‡, at low pressures of the order of 1 millimetre. He observed the continuous spectrum as a faint background when the discharge took place in a quartz capillary tube, and not at all when the discharge occurred in glass tubes. He suggested that the continuous spectrum in krypton might be due to fluorescence of the quartz. It is possible that the continuous spectrum which he observed is different from those which have been observed in this laboratory in monatomic gases. These spectra become prominent in gases at high pressures in wide tubes, and would not be noticeable in gases at pressures of the order of 1 millimetre in capillary tubes. The continuous spectra in argon, helium, and neon do not originate from fluorescence of the quartz, since we have observed them equally well in quartz, pyrex, and glass tubes§.

In the high-frequency discharges in argon at pressures from 5 mm. to 150 mm. in tubes 2 cm. and 3 cm. diameter the argon red spectrum only is observed in addition to the continuous spectrum. This is to be expected, since the mean energy of the electrons is small, from 12 volts to 7 volts, in this range of pressures, while the excitation potential of the argon blue spectrum

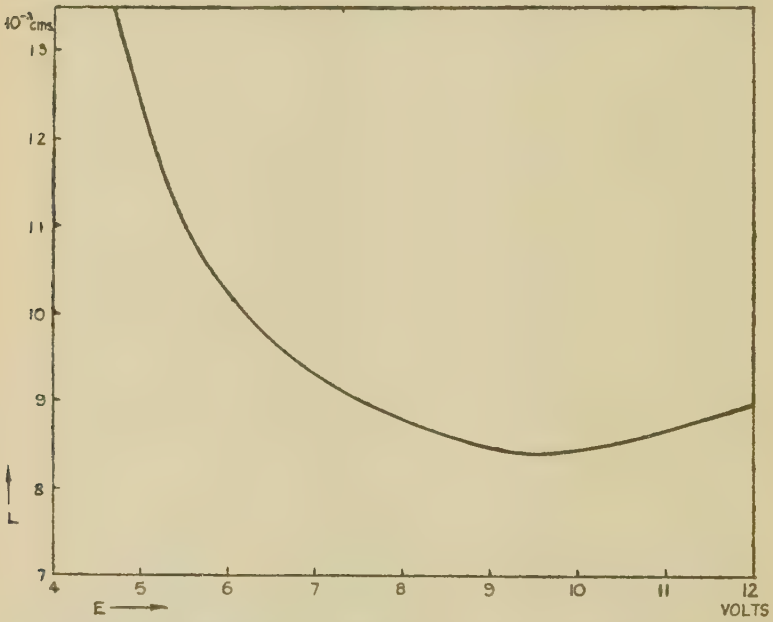
* J. S. Townsend and M. H. Pakkala, Phil. Mag. xiv. p. 418 (1932).

† P. Johnson, Phil. Mag. x. p. 921 (1930).

‡ W. F. Meggers, T. L. de Bruin, & C. J. Humphreys, 'Bureau of Standards, Journal of Research,' iii. p. 129 (1929).

§ Since writing the above, we find that Dorgelo and Washington (Proc. Amsterdam, xxx. p. 33 (1927)) have described a continuous spectrum in argon which was observed by Penning, and they suggested that it might be due to a slight admixture of hydrogen. In our experiments when slight traces of hydrogen were present it appeared in the spectrum in the form of lines superimposed upon the continuous background, and when these last traces of hydrogen were removed by means of palladium black and cold calcium the continuous spectrum remained unchanged while the lines due to the hydrogen disappeared.

is about 34 volts. At the higher pressures, however, even the lines of the red spectrum are extremely faint. In the unstriated discharge at 150 mm. pressure the only line of appreciable intensity appearing in the whole range from 7200 Å. to 2200 Å. is the line at 6965 Å. ($1s_5$–$2p_2$), whose excitation potential is about 13·3 volts. At this pressure, where the mean energy of agitation of the electrons is 6 volts, the principal blue and violet lines corresponding to the next highest excited level at 15·7 volts are absent. This affords a striking contrast with the case of helium,



Relation between the mean free path L of electrons in argon and their energy E.

where lines having excitation potentials up to 23 volts are intensely excited when the mean energy of the electrons is only 3 or 4 volts. It can only be concluded from this that the distribution of the energies of the electrons in argon about the mean energy is much narrower than the distribution in helium.

This narrow distribution of the energies of the electrons may be explained by considering the variation of the mean free paths of the electrons with their velocity *.

* The values for this curve are taken from 'The Motion of Electrons in Gases,' *l. c.*

The mean free path, L, of the electrons in argon at 1 millimetre pressure is shown in the figure on p. 196 in terms of the energy of agitation E* for values of E from 4 to 13 volts It will be seen that for values of $\overline{E}$ from 4 volts to 9 volts. the value of L decreases rapidly as E increases. Thus electrons whose energy E is less than the mean E will make considerably fewer collisions than the average, and their velocity in the direction of the electric field will be much greater than the mean, so that they will gain energy from the field. The probability of electrons having energies much smaller than the mean is thus greatly reduced. Similarly electrons whose energy E is greater than the mean will have a shorter free path and will consequently lose more energy in collisions, so that the probability of electrons having energies much greater than the mean is also considerably reduced. For values of E greater than ten volts the distribution is likely to be wider, since in this region the mean free path increases as the energy of the electrons is raised. In general the continuous spectrum is intense when the mean energy of the electrons is less than about nine volts, and since the distribution here is narrow very few of the lines are excited with appreciable intensity. When the mean energy exceeds 10 volts the distribution becomes wider and many electrons have sufficient energy to excite the line spectrum.

We wish to thank Prof. Townsend for much helpful advice and criticism.

XIV. *Notices respecting New Books.*

An Introduction to the Theory of Canonical Matrices. By H. W. Turnbull, M.A., F.R.S., and A. C. Aitken, D.Sc. [Pp. xiii+192.] (Blackie & Son, Ltd., 1932. Price 17*s.* 6*d.*)

IT is largely due to the work of Professor Turnbull that interest in algebra has revived in England, and he has made the modern developments of certain of its branches more accessible by means of his former book on the theory of invariants. He has now, in collaboration with Dr. Aitken, performed the same service for some of the most important parts of the theory of matrices. His earlier book gave an account of their fundamental properties, but these are again summarised in

* J. S. Townsend, Proc. Roy. Soc. A, cxx. p. 516 (1928).

the first chapters of the work under review, so that it is self-contained, and may be read by anyone who is familiar with the elementary properties of determinants.

Any bilinear form in two sets of variables u and x can be written in matrix notation as uAx, where A is the matrix of the form. If u and x are subjected to separate linear transformations $u=v\mathrm{P}$, $x=\mathrm{Q}\xi$, the new form is $v\mathrm{B}\xi$, where $\mathrm{B}=\mathrm{PAQ}$. Thus transformations of the variables lead to transformations of the matrix A, and the reduction of the form to a standard (or *canonical*) shape depends on a suitable determination of the matrices P and Q. In dealing with applications, the transformation involved will normally belong to a particular sub-group of the general group of equivalent matrices. Thus, in the matrix form associated with the dynamical theory of small oscillations, the variables are congruent, and there is a particular relation between P and Q. The book discusses with great thoroughness and originality of treatment the reduction to canonical form in the case of all the important sub-groups, showing, *inter alia*, its dependence on the characteristic equation and invariant factors of the matrix. The final chapter gives the application of these results to problems of analysis, dynamics, and statistics.

The book is essential to all students of algebra, and should interest also those applied mathematicians and physicists who like to know the mathematical basis for the results they use. The authors do not deal with infinite matrices, which enter so vitally into the theory of integral equations and quantum theory. May we hope that they will be discussed in a subsequent volume ?

A Comprehensive Treatise on Inorganic and Theoretical Chemistry.—Vol. XII. By J. W. Mellor, D.Sc., F.R.S. [Pp. xii+944.] (Longmans, Green & Co., Ltd. Price 63*s*. net.)

The twelfth volume of Dr. Mellor's monumental Treatise has, perhaps, a peculiar interest for physicists, in that it includes not only uranium, the compounds of which are so important for radioactivity and for the study of phosphorescence, but also masurium and rhenium, the new elements discovered by Noddack and Tacke at the instigation of W. Hausser, by means of X-Ray analysis. There are already 17 pages devoted to the chemistry of these new elements.

About half the volume is devoted to iron, the bibliography of which is so appalling that one can only marvel at the courage displayed by Dr. Mellor and his army of collaborators in their task. Even to copy down the thirty closely printed pages of reference to but one section, that on the *Occurrence of Iron*, let alone to digest their contents, might well dismay the less

hardened. We trust that the attribution "Old Tale" of four lines of Kipling quoted is not an index to the reliability of the reference. The mass of material is imposing, but it is to be remarked, however, that while the physical properties of, for instance, manganese and other elements, are discussed early in the articles on these elements, including even (although perhaps in too condensed a form to be very useful) the spectrum and the hyperfine structure of its lines, there is practically nothing of the physical properties of iron, either spectral or magnetic. Probably the physical properties are reserved for a further treatment of iron, to appear later, for on the half-title we find "Fe (Part I.)" although nowhere else—not in the chapter headings, for instance—do we find any specific indication that there is more on iron still to come. The crystalline structure is discussed at much length.

The volume will maintain the reputation which the Treatise has won as a reference book.

Great Men of Science. By PHILIPP LENARD. Translated from the second German edition by H. STAFFORD HATFIELD. With a preface by Prof. E. N. da C. ANDRADE. (London: G. Bell & Sons. Price 12*s.* 6*d.* net.)

THIS is the work of a great man, a pioneer in physics. It is a story of great men, who were themselves pioneers, starting with Pythagoras of Samos, Euclid, and Archimedes, and ending with Hasenöhrl and his speculations into the connexions between energy and mass. It is the outcome of Lenard's own historical research into the life-history of great men of science. He found that these scientists were often much more above the common run of humanity than the most widely read biographies suggested. His "joy was great to find that these personalities so well matched the greatness of their achievements that they were fit to serve as examples to future generations both from the point of view of their work and from that of their lives."

Lenard has brought great learning and powers of criticism to his task. The book is singularly free from national prejudices. He is as much at home in discussing the work of Newton, Black, Watt, Priestley, Cavendish, Rumford, Dalton, Davy, Faraday, Joule, Kelvin, Darwin, Maxwell, and Crookes as in that of Ohm, Gauss, Weber, Mayer, Helmholtz, Clausius, Bunsen, Kirchhoff, Boltzmann, and Hertz. He is astonishingly "at home" in his knowledge of the personal life of those with whom he deals, and he knows how to select those incidents which bring a picture of the character of the man before us. His description of the work of James Watt is an excellent example of his power in portrayal. The following sentence is also illustrative:—"Tycho (Brahe) had reddish hair and was a great friend of animals."

It is impossible to do more than hint in this way as to the importance of this book. It should be in the hands of every student of physics who has been through an introductory junior course in which he shall have acquired a general knowledge of what physics is all about.

It is illustrated by sixty-two portraits (mostly full-paged) of those of whom he writes. These add greatly to the value of the book.

The translation appears to have been well done on the whole. Now and then a difficulty has clearly been felt in adapting German to English construction. Further, on p. 126, line 7 from bottom, "independent" should be "dependent."

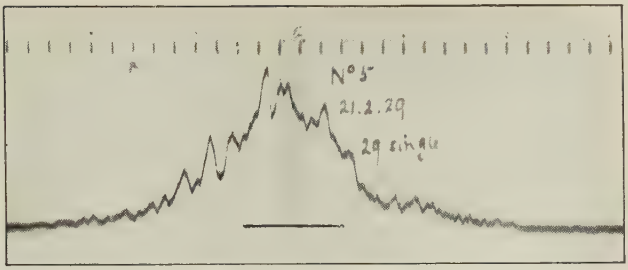
Principles of Descriptive Geometry. By E. L. Ince, M.A., D.Sc. [Pp. viii+152.] (London: Edward Arnold & Co., 1933. Price 8*s.* 6*d.*)

Dr. Ince points out in the preface to this little book that many of the existing text-books, particularly those in the English language, deal with the methods rather than with the principles of the subject. Perhaps this is due to the fact that most of the textbooks written for students in their last years at school or in their first year at a university are devoted mainly to other aspects of solid geometry, and descriptive geometry, if treated at all, is treated very briefly. Dr. Ince, not being distracted by the many other aspects of solid geometry, has produced a well-balanced and very readable little book on descriptive geometry, which he aptly defines as "the technique of representing a solid object as a plane figure, not merely pictorially, but accurately and with complete definiteness."

We should expect a book on such a subject to be well equipped with diagrams and we are not disappointed, for there are 153 well-drawn diagrams in 152 pages. The subject-matter is restricted to solids with planes faces; cylinders, cones, surfaces of revolution, etc., are to be treated in a later volume, the appearance of which depends on the reception given to the present volume. The book is divided into seven chapters:—an introduction, three chapters on the point, the straight line, and the plane in biorthogonal projection, chapters on displacements, on the representation of various polyhedra, and on the figured plan. At the end of each chapter there is a set of examples, and, finally, a set of practical exercises at the end of the book. The book is attractively written and well produced; we hope that the second volume will soon appear.

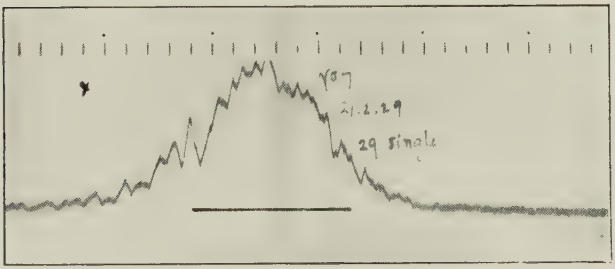
[*The Editors do not hold themselves responsible for the views expressed by their correspondents.*]

FIG. 2.



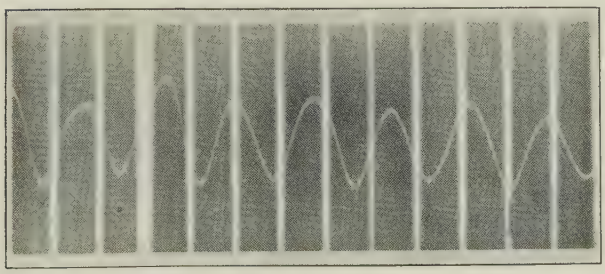
Record of fundamental airscrew sound by hot-wire microphone and Moll galvanometer. The markings at the top indicate seconds. Height of aeroplane 1000 ft.

FIG. 3.



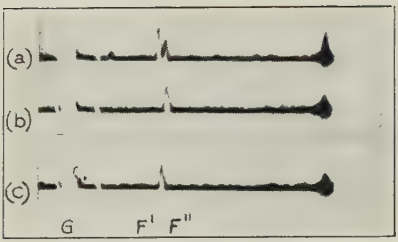
Record of fundamental airscrew sound by hot-wire microphone and Moll galvanometer. Height of aeroplane 1000 ft.

FIG. 4.



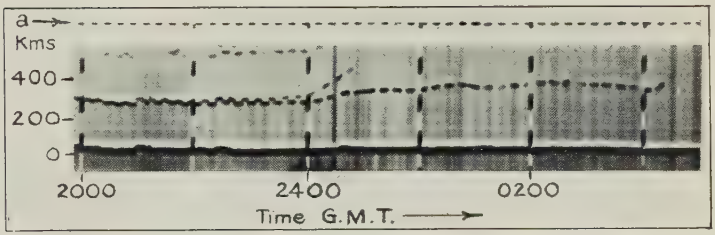
Record of fundamental airscrew sound by hot-wire microphone and Einthoven galvanometer. The time-lines divide the record into hundredths of a second.

Fig. 4.



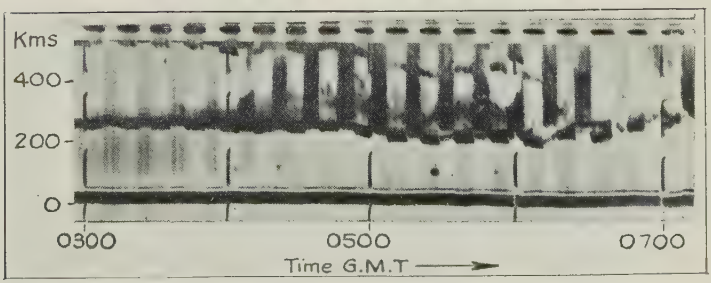
Polarization of double F echo.

Fig. 5.



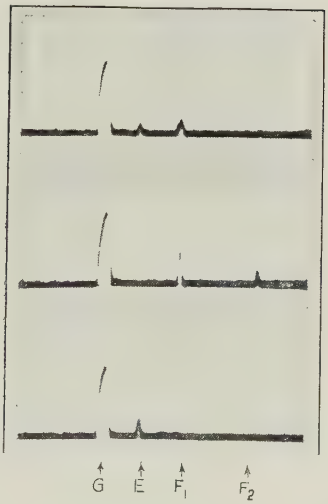
Polarization of split F echo.

Fig. 6



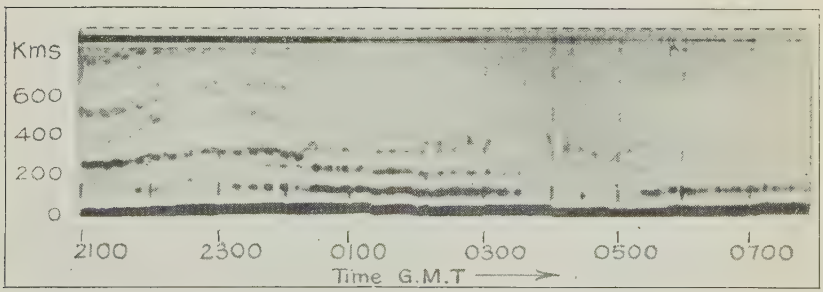
Polarization of broadened F echo.

Fig. 7.



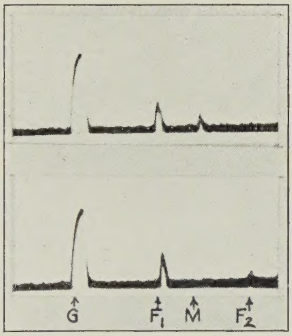
Polarization of E and F echoes during transition period.

Fig. 8.



Multiple echoes.

FIG. 10.

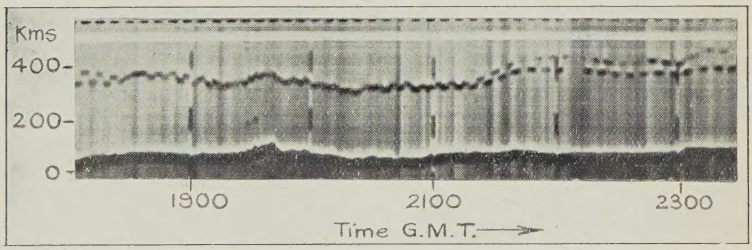


Polarized Receiver (R).

Polarized Receiver (L).

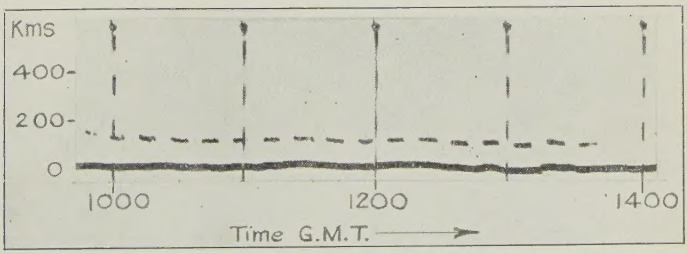
Polarization of M echo.

FIG. 12.



Reversed splitting.

FIG. 13.



Daytime polarization.

On this record the lamp indicator was not in action. Visual observation showed that the received signal had left-handed polarization.

High frequency discharges in argon.